شناسایی آماری الگو

* مشخصات استاد

. دکتر رحمتی

. آدرس پست الکترونیکی rahmati@aut.ac.ir

. ارتباط با استاد

to: rahmati@aut.ac.ir

subject: SPR member

content:

Name: _____

S.N: _____

Email: _____

* منابع

1. Introduction To Statistical Pattern Recognition
K. Fukunaga
Academic Press, 1990
2. Pattern Recognition
S. Theodoridis & K. Koutroumbass
Academis Press, 2003
3. Pattern Classification

* نگارش

. آرش مشکی

. نیمسال اول تحصیلی ۸۸ - ۸۷

. ☎ ۰۹۱۲- ۲۱۲۹۴۰۴

* ارزیابی

- تکالیف (20%) ← پروژه‌های کامپیوتری + حل مسائل
- آزمون‌ها (60-65 %) ← میان‌ترم + پایان‌ترم
- پروژه پایانی (16-20 % ← گزارش + ارائه + تحقیق ورای پیاده سازی

* مباحث

- مقدمات
  - آمار و احتمالات
  - جبر خطی
- تشخیص آماری الگو
- کاهش ابعاد
  - PCA
- خوشه‌بندی

* کاربردها
- بینایی ماشین
  - خط تولید: صحت قطعات، تشخیص قطعات
  - تشخیص هدف خودکار: نظامی، پزشکی
  - تصاویر هوایی و ماهواره‌ای
- شناسایی حروف
  - تشخیص مقصد نامه‌های پستی
  - خواندن متن اسکن شده
- پزشکی
  - کنترل EEG، ECG
  - تصاویر پزشکی
  - کمک به پزشک
- کنترل
  - HCI (Human-Comp. Interface)
  - BCI (Brain-Comp. Interface)
- Action Recognition
  - سیستم‌های نظارتی
  - Video Sequencing

* اجزای سیستم
- دنیای واقعی
- اخذ داده‌ها (سنسورها) ← DBها، دوربینها، ECG، ...
- پیش پردازش ← حذف نویز، نرمال سازی، استخراج ویژگیها، ...
- کاهش ابعاد ← حذف ویژگیها، ادغام آنها، ...
- پیش بینی ← کلاس بندی، خوشه بندی، ...
- انتخاب مدل

* انواع پیش بینی
  - کلاسه بندی
  - طبقه بندی
  - Regression
  - Desciption

* Feature
  - معرف ویژگیهای یک شی
  - یک شی را با برداری از ویژگیها نشان می دهند.
  - انتخاب ویژگیها در کلاسه بندی موثر و بسیار مهم است
  - ویژگیها دارای مقادیر جدا هستند.
  - شی های کلاسهای مختلف باید دارای مقادیر مجزا باشند.

* الگو
  - ترکیبی از ویژگیها
  - الگو = مجموعه ای از ویژگیها + معانی آنها

* طبقه بندی
  - فضای ویژگیها را به نواحی مختلف افراز می کنند.
  - مرز بین نواحی را dicision boundry می گویند.
  - تابع مدل این مرز هدف الگوریتم های کلاسه بندی است.

* روشهای شناسایی الگو
  - آماری ← متغیرهای تصادفی
  - شبکه عصبی ← شبکه ای از سلولهای پردازنده
  - ساختاری ← زیرساختارها و تطبیق ارتباط ها
  - درخت تصمیم ← ؟

* هر ویژگی، یک متغیر تصادفی است

- به‌ازای هر مقدار آن، یک فراوانی موجود است.
- مقادیر مختلفی را می‌پذیرد
- چند ویژگی را به صورت بردار نشان می‌دهند (بردار ستونی)

$$\underbrace{\underline{X}}_{n\times 1 \text{ بردار}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1 \quad \dots \quad X_n]^T$$

- دارای یک تابع چگالی احتمال نشان داده می‌شود

$p(x)$: probability density function

- دارای یک تابع توزیع تجمعی است

$P(x)$: probability distribution function $= \text{Prob}\{X \leq x\}$

- علائم

X : Random Var.
$\underline{X}$: Random vector
x : Deterministic Var.
d, n : Dimension
T : Transpose

$[\Sigma]$: Covarience Matrix of $\underline{X}$
N : # of samples
C, M : # of classes
ω_i : label of class $i = 1, \dots, M$

- روابط

$* \; P(\underline{x}) \equiv P(x_1, \dots, x_n) = \text{Prob}\{X_1 \leq x_1, \dots, X_n \leq x_n\} = \text{Prob}\{\underline{X} \leq \underline{x}\}$

Joint Distribution

$$* \; \frac{\partial P_x(\underline{x})}{\partial \underline{x}} = p_{\underline{x}}(\underline{x}) = p(\underline{x}) = \frac{\partial^n P(\underline{x})}{\partial x_1 \, \partial x_2 \dots \partial x_n}$$

$$* \; P(\underline{x}) = \underbrace{\int_{-\infty}^{+\infty}}_{n \text{ تایی}} p(\underline{x}) \, d\underline{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{+\infty} p(x_1 \dots x_n) \, dx_1 \dots dx_n$$

* $\omega_i \qquad i = 1, \dots, M$

$\underline{x} \in \omega_i \longrightarrow \underline{x}$ belongs to i^{th} class

$p(\underline{x} | \omega_i) \equiv p_i(\underline{x})$

conditional density function of class i

* $P_{\underline{X}}(\underline{x}) \equiv p(\underline{x}) = \sum_{i=1}^{M} P_i \cdot p_i(\underline{x})$

mixture (total) density function of $\underline{X}$

$P_i = \text{Prob}\{\omega_i\}$

a priori probability of ω_i (احتمال اولیه)

* $P(\omega_i | \underline{x}) \equiv q_i(\underline{x})$

posterior probability of ω_i (احتمال ثانویه)

* $P(\omega_i , \underline{x}) = P_i \cdot p_i(\underline{x}) = P(\omega_i | \underline{x}) \cdot p(\underline{x})$

$$q_i(\underline{x}) = P(\omega_i | x) = \frac{P_i \cdot P_i(x)}{P(\underline{x})}$$ $\leftarrow$ نرمالیزه کردن

* $i = 1, 2$

$$q_1(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} q_x(\underline{x})$$

$$P_1 \cdot p_1(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P_2 \cdot P_2(\underline{x})$$

$\ell(\underline{x})$ ... likelihood $= \dfrac{P_1(x)}{P_2(\underline{x})}$

* $m = E\{X\}$

$\sigma_X^2 = E\{(X - m)^2\}$

• معرفی پارامترهای توابع توزیع

$P_i(\underline{X}) \qquad i = 1, \dots, M$

* میانگین یک بعدی

$$\mu = m = E\{X\} = \int_{-\infty}^{+\infty} x \cdot p(x)\, dx$$

* میانگین چندبعدی

$$\underline{\mu} = \underline{m} = E\{\underline{X} \mid \omega_i\} = \int_{-\infty}^{+\infty} \underline{x} \cdot p(x) \cdot d\underline{x}$$

i^{th} element of $\underline{X}$ *

$$m^i = i^{th} \text{ element of } E\{\underline{X}\} = \int_{-\infty}^{+\infty} x_i \cdot p(x_i) = \underbrace{\int_{-\infty}^{+\infty}}_{\text{بعد } n} x_i \cdot p(x) \cdot d\underline{x}$$

marginal density

$$p(x_i) = \underbrace{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty}}_{\text{بعد } n-1} p(\underline{x})\, dx_1 \dots dx_{i-1}\, dx_{i+1} \dots dx_n$$

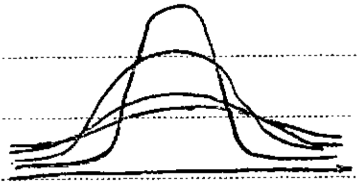
* ماتریس کوواریانس

$$[C] = [\Sigma]_x = E\{(\underline{X} - \underline{m})(\underline{X} - \underline{m})^T\}$$

$$= \begin{bmatrix} E\{(X_1 - m_1)^2\} = \sigma_1^2 & & & \\ E\{(X_2 - m_2)(X_1 - m_1)\} & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$$

$$c_{ij} = E\{(X_i - m_i)(X_j - m_j)\}$$

$i, j = 1, \dots, n$



$C_{ii} = \sigma_i^2$ *

$C_{ij} = \rho_{ij} \cdot \sigma_i \cdot \sigma_j$

$\rho_{ij} \equiv$ correlation coefficient of X_i & X_j

$\sigma_i \equiv$ Standard Deviation of X_i

*

$$[C] = [\Sigma]_X = E\{(\underline{X}-\underline{m})(\underline{X}-\underline{m})^T\}$$

$$= E\{\underline{X}X^T - \underline{X}\underline{m}^T - \underline{m}X^T + \underline{m}m^T\}$$

$$= \underbrace{E\{\underline{X}\,\underline{X}^T\}} + \underline{m}m^T$$

$[R] = [\Sigma] \equiv$ Convolution Matrix of $\underline{X}$

*

$\underline{m}_{n\times 1} \equiv$ mean vector $= E\{\underline{X}\}$

$$[\Sigma]_{n\times n} \equiv E\{(\underline{X}-\underline{m})(\underline{X}-\underline{m})^T\} = \{C_{ij} \mid C_{ij} = \rho_{ij}\cdot\sigma_i\cdot\sigma_j\}$$

$$[\Sigma] = [T][R][T]$$

$$[T] = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \emptyset \\ \emptyset & & \\ & & \sigma_n \end{bmatrix} \qquad [R] = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ . & 1 & & & \\ & & & & \\ \rho_{n} & & & & 1 \end{bmatrix}$$

$\rho_{ij} \equiv$ correlation coeff. of x_i and x_j

(Gausian Dist) Normal Distribution *

$$\underset{x}{P}(\underline{x}) = N(\underline{m}, [\Sigma]) \quad •$$

$$= \frac{1}{(2\pi)^{n/2} \cdot |[\Sigma]|^{1/2}} \exp\{-\tfrac{1}{2} d(\underline{x})\}$$

•

$d(\underline{x}) \equiv$ distance function $\equiv$ Mahanobis dist $= (\underline{x} - \underline{m})^T [\Sigma]^{-1} (\underline{x} - \underline{m})$

$|\cdot| \equiv$ determinant of $[\Sigma]$

• خواص توزیع نرمال

Physical Justification -

معرفی توزیع $\longleftarrow$ $\underline{m}$, $[\Sigma]$ -

independent $\Longleftarrow$ uncorrelated -

uncorrelated $\Longleftrightarrow$ $c_{ij} = 0 \quad , i \neq j$

$$|\Sigma| = \begin{bmatrix} \sigma^2 & & \emptyset \\ & \ddots & \\ \emptyset & & \sigma^2 \end{bmatrix}$$

uncorrelated $\Longleftrightarrow$ $[\Sigma] = [\Sigma]^{-1}$ ~ $d(\underline{x})$ میزان تغییر

$p(\underline{x}) = p(x_1, x_2, \dots, x_n) = p(x_1) \cdot p(x_2) \dots p(x_n) \Rightarrow$ independent

Normal Marginal Distribution -

Normal Conditional Distribution -

Linear Transformation -

$\underline{Y} = [a]^T \underline{X}$

$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}_{2\times1}$ $\underline{Y} = [Y_1]_{1\times1}$ ☑

$$P_{\underline{X}\underline{Y}}(x_1, x_2, y_1) = \begin{cases} (x_1 + 3x_2)\, y_1 & , \ 0 \leqslant x_1, x_2, x_3 \leqslant 1 \\ 0 & , \ \text{otherwise} \end{cases}$$

- event $\underline{Y} \leqslant \frac{1}{2}$
  $\text{Prob}\{\underline{Y} \leqslant \frac{1}{2}\} = \int_0^1 \int_{-\infty}^{1/2} (x_1 + 3x_2)\, y_1 \cdot d\underline{x} \ d\underline{y} = 1/4$

- marginal dist $\underline{X}$
  $P_{\underline{X}}(\underline{x}) = \int_{-\infty}^{+\infty} P_{\underline{X}\underline{Y}}(\underline{x}, \underline{y}) \cdot d\underline{y} = \int_0^1 (x_1 + 3x_2) \cdot y_1 \cdot dy_1$

$$= \begin{cases} 0 & , \text{otherwise} \\ \frac{1}{2}(x_1 + 3x_2) & , 0 \leqslant x_1, x_2 \leqslant 1 \end{cases}$$

- $\underline{X}$ & $\underline{Y}$ independent ?
  $P_{\underline{X}\underline{Y}}(\underline{x}, \underline{y}) \overset{?}{=} P_{\underline{X}}(\underline{x}) \cdot P_{\underline{Y}}(\underline{y})$

* خواص $[\Sigma]$

- Symmetric — متقارن
- Positive Definite — مثبت معین

$$\underline{a}^T \underset{n\times n}{[\Sigma]} \underline{a} > 0$$

$\underline{a} \longrightarrow$ nx1 . هر بردار

- $|\Sigma| > 0$
- $[\Sigma]^{-1}$ exist.
- eigenvalues of $[\Sigma] > 0$ $\xrightarrow{\underline{\text{[illegible]}}}$ متمایز

$$[\Sigma]_{n\times n}\, \underline{x} = \underline{\lambda}\, \underline{x}_{n\times 1}$$

$$\underline{\lambda} = cte \xrightarrow{[\Sigma] > 0} \lambda_1, \lambda_2, \dots, \lambda_n > 0$$

positive semi-definite *

$$\underline{a}^T [\Sigma] \cdot \underline{a} \geqslant 0$$

☑ مثبت قطعی

$$[A] = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

* تخمین زدن پارامترهای $\underline{m}$

1-D

$X : x_i \quad i = 1, \dots, N$

$E\{X\} = m$

$\hat{m}_1 = \frac{1}{N} \sum_{i=1}^{N} x_i$ ① N تا نمونه

$\hat{m}_2$ ② N تا نمونه

⋮

$\hat{M}$ is Random Variable ↲

$$E\{\hat{M}\} = \frac{1}{N} E\{\sum_{i=1}^{N} X_i\} = \frac{1}{N} \sum_{i=1}^{N} E\{X_i\} = \frac{1}{N} \cdot N \cdot m = m$$

Unbiased Estimation : اگر حدس ما طوری باشد که رابطه‌ی بالا برقرار به m برسد

$$\hat{\sigma}_M^2 = Var\{\hat{M}\} = E\{(\hat{M}-m)^2\}$$

$$= E\{(\frac{1}{N}\sum_{i=1}^{N} X_i - m)^2\}$$

$$= E\{(\frac{1}{N}\sum_{i=1}^{N}(X_i-m))(\frac{1}{N}\sum_{j=1}^{N}(X_j-m))\}$$

با فرض استقلال آماری $= E\{(X_i-m)(X_j-m)\}$

$$= E\{(X_i-m)\} . E\{(X_j-m)\}$$

$$= \frac{1}{N^2} . N . \sigma^2 = \frac{1}{N}\sigma^2$$

$$N \longrightarrow \infty \ : \ \hat{\sigma}_M^2 = 0$$

Consistant Estimation : با بالا بردن تعداد نمونه ، واریانس به طرف صفر میل کند.

n-D •

$$\underline{m} = E\{\underline{X}\} = \frac{1}{N}\sum_{i=1}^{N}\underline{X}_i$$

* تخمین زدن پارامتر [Σ]

1-D •

X

$m = E\{X\}$ ⟶ تخمین زده می‌شود

$\sigma^2 = E\{(X-m)^2\}$

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

$$E\{\hat{\sigma}^2\} = \frac{1}{N-1}\sum_{i=1}^{N} E\{X_i - \bar{X}\}^2 \qquad \pm m$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}(\sigma^2 + \frac{1}{N}\sigma^2 - 2\overset{1/N\,\sigma^2}{E\{(X_i-m)(\bar{X}-m)\}})$$

$$= \sigma^2$$

n-D •

$\underline{X}$

$$\text{cov}\{\underline{X}\} = [\hat{\Sigma}] = \frac{1}{N-1} \Sigma_{i=1}^{N} \left((\underline{x}_i - \underline{m})(\underline{x}_i - \underline{m})^T \right)$$

• راه آسانتر

$$[\Sigma] = [S] - \underline{m}\,\underline{m}^T$$

↳ autocorrelation matrix

$$[\hat{S}] = \frac{1}{N} \Sigma_{i=1}^{N} \underline{X}_i \cdot \underline{X}_i^T$$

* Online Learning
Iterative Learning
Recursive Learning
Incremental Learning

وقتی که N نمونه در دسترس نباشد به تدریج واردی شود

$\underline{m}_N \equiv$ mean of $\underline{X}$ based on first N samples

$$\underline{m}_N = \frac{1}{N} \Sigma_{i=1}^{N} \underline{x}_i$$
$$= \frac{1}{N} \left[\Sigma_{i=1}^{N-1} \underline{x}_i + \underline{x}_N \right]$$
$$= \frac{1}{N} \left(\frac{N-1}{N-1} \right) \Sigma_{i=1}^{N-1} \underline{x}_i + \frac{1}{N} \underline{x}_N$$

$$= \frac{N-1}{N} \cdot \frac{1}{N-1} \Sigma_{i=1}^{N-1} \underline{x}_i + \frac{1}{N} \underline{x}_N$$
$$= \frac{N-1}{N} \underline{m}_{N-1} + \frac{1}{N} \underline{x}_N$$
$$= \underline{m}_{N-1} - \frac{1}{N} \underline{m}_{N-1} + \frac{1}{N} x_N$$

$$\underline{m}_N = m_{N-1} + \frac{1}{N} \left(\underline{x}_N - \underline{m}_{N-1} \right)$$ تاثیر نمونه‌های آخر ضعیف تر می‌شود

$$\underline{m}_N = \underline{m}_{N-1} + \alpha (\underline{x}_N - \underline{m}_{N-1})$$ بهنگام می‌شود.

ضریب $\leftarrow \frac{1}{N}$: برای حالتی که متغیر دارای خاصیت Stationary باشد

$\swarrow \; \alpha$:

$$\boxed{\underline{m}_N = (1-\alpha)\, \underline{m}_{N-1} + \alpha\, \underline{x}_N}$$

$$[S] = \frac{1}{N} \sum_{i=1}^{N} \underline{X}_i \cdot \underline{X}_i^T$$

$$\boxed{[S]_N = (1-\alpha)\, [S]_{N-1} + \alpha\, \underline{X}_N \cdot \underline{X}_N^T}$$

$$[\Sigma] = [S] - \underline{m}\, \underline{m}^T$$

$$[\Sigma]_N = \left[(1-\alpha)[S]_{N-1} + \alpha\, \underline{X}_N\, \underline{X}_N^T\right] - \left[\sim\right]\left[\sim\right]^T$$

$$\boxed{[\Sigma]_N = (1-\alpha)\, [\Sigma]_{N-1} + \alpha (\underline{X}_N - \underline{m}_{N-1})(\underline{X}_N - \underline{m}_{N-1})^T}$$ OK

$A = n \times n \qquad C = m \times m \qquad B, D = n \times m$

$$[A + BCD^T]^{-1} = A^{-1} - A^{-1} B \left[C^{-1} + D^T A^{-1} B \right] D^T A^{-1}$$

$$[S]_N = \overbrace{(1-\alpha)\, [S]_{N-1}}^{A} + \overbrace{\alpha}^{C}\, \overbrace{\underline{X}_N}^{B}\, \overbrace{\underline{X}_N^T}^{D}$$

$$[S]_N^{-1} = \frac{1}{1-\alpha} \left([S]_{N-1}^{-1} - \alpha \cdot \frac{[S]_{N-1}^{-1}\; X_N \cdot \underline{X}_N^T \cdot [S]_{N-1}^{-1}}{1 + \alpha\, (\underline{X}_N^T \cdot [\sim]_{N-1}^{-1} \cdot \underline{X}_N - 1)} \right)$$

$I.$ $[S]^{-1}_{0} \equiv$ initial inverse matrix $[S]^{-1}$

$II.$ $\underline{w} = [S]^{-1}_{0} \cdot \underline{X}_1$

$III.$ Update Equation

$$[S]^{-1}_{K} = \frac{1}{1-\alpha}\left([S]^{-1}_{K-1} - \alpha \frac{\underline{w}\ \underline{w}^T}{1+\alpha(\underline{w}^T \underline{X}_K - 1)}\right)$$

- استفاده از تخمین $\hat{\underline{M}}$

$$[\hat{\Sigma}] = \frac{1}{N}\sum_{i=1}^{N} (\underline{X}_i - \hat{\underline{M}})(\underline{X}_i - \hat{\underline{M}})^T$$

آزمون unbiased بودن

$$[\hat{\Sigma}] = \frac{1}{N}\sum_{i=1}^{N} \{(\underline{X}_i - \underline{m}) - (\hat{\underline{M}} - \underline{m})\}\{(\underline{X}_i - \underline{m}) - (\hat{\underline{M}} - \underline{m})\}^T$$

$$= \frac{1}{N}\sum_{i=1}^{N} (\underline{X}_i - \underline{m})(\underline{X}_i - \underline{m})^T - (\hat{\underline{M}} - \hat{\underline{m}})(\hat{\underline{M}} - \hat{\underline{m}})^T$$

با برداشتن امید ریاضی

$$E\{[\hat{\Sigma}]\} = [\Sigma] - \frac{1}{N}[\Sigma] = \frac{N-1}{N}[\Sigma] \rightarrow \text{biased!}$$

پس برای unbiased کردن آن فرمول را تغییر می‌دهیم.

$$[\hat{\Sigma}] = \frac{1}{N-1} \sum (\underline{x}_i - \underline{m})(\underline{x}_i - \underline{m})^T$$

* تبدیل خطی

- هدف : کاهش ابعاد بردارها / بدست آوردن یک خاصیت در فضای جدید
uncorrelated ←

-

$$\underline{Y}_{m\times 1} = [a]^T \underline{X}_{n\times 1}$$

$[a] \equiv n\times m$

$$\underline{Y} = [a]^T \underline{X}$$

$$E\{\underline{Y}\} = \underline{m}_Y = [a]^T E\{\underline{X}\}$$

$$\underline{m}_y = [a]^T \underline{m}_x$$

$$\mathrm{cov}\{\underline{Y}\} \equiv [\Sigma]_y = E\{(\underline{Y} - \underline{m}_y)(\underline{Y} - \underline{m}_y)^T\}$$

$$= E\{[a]^T(\underline{X} - \underline{m}_x)(\underline{X} - \underline{m}_x)^T [a]\}$$

$$= [a]^T \; E\{\sim\} \; [a]$$

$$[\Sigma]_y = [a]^T [\Sigma]_x [a]$$

$$[S]_y = E\{\underline{Y}\underline{Y}^T\} = [a]^T [S]_x [a]$$

$$\underline{X} : N(\underline{m}_x, [\Sigma]_x)$$

$$p(\underline{x}) = \frac{1}{(2\pi)^{n/2} |[\Sigma]|^{1/2}} \exp\{-\tfrac{1}{2} d^2(\underline{x})\}$$

$$d^2(\underline{x}) = (\underline{x} - \underline{m}_x)^T [\Sigma]_x^{-1} (\underline{x} - \underline{m}_x)$$

$$p(\underline{y}) = \frac{1}{|[a]| \; (2\pi)^{n/2} |[\Sigma]_x|^{1/2}} \exp\{-\tfrac{1}{2} d^2(\underline{y})\}$$

$$|\Sigma_y| = |[\Sigma]_x| \times |[a]^T|^2$$

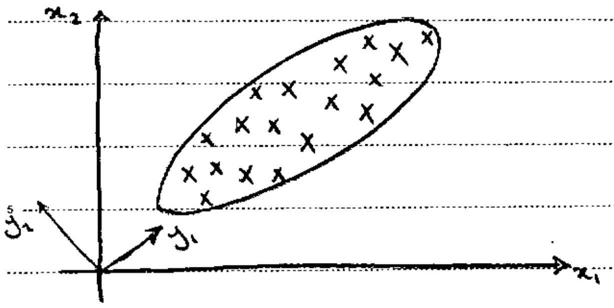
$$d_y^2(y) = (\underline{y} - \underline{m}_y)^T [\Sigma]_y^{-1} (\underline{y} - \underline{m}_y)$$

$$= (x - \underline{m}_x)^T [a][a]^{-1} [\Sigma]_x^{-1} ([a]^T)^{-1} [a]^T (\underline{x} - \underline{m}_x)$$

$$= (\underline{x} - \underline{m}_x)^T [\Sigma]_x^{-1} (\underline{x} - \underline{m}_x)$$

under linear transformation, distance is preserved.

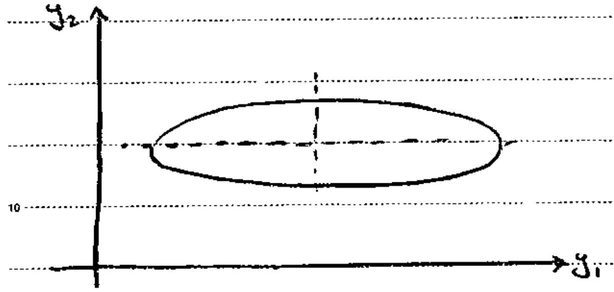
X - Space



→ Normal Dist

→ x_1 & x_2 are correlated

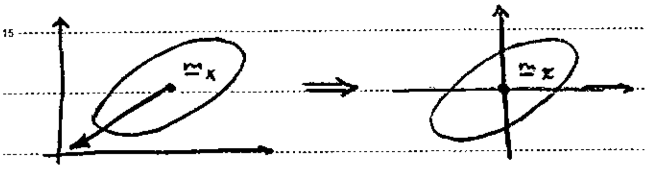
linear transformation



→ y_1 & y_2 are not correlated

وقتی محورهای بیضی موازی محورها باشند.

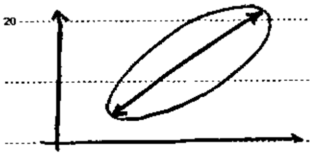
Orthonormal Transformation



$$\underline{z} = \underline{x} - \underline{m}_x$$

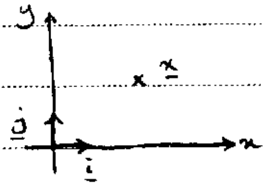
$$d_{\underline{z}}^2 = \underline{z}^T [\Sigma]_x^{-1} \underline{z}$$

• می‌خواهیم بیشترین مقدار فاصله را به کمک تابع تغییری بیابیم.
با این قید:



constraint : $\underline{z}^T \underline{z} = 1$

$$F = \underline{z}^T [\Sigma]_x^{-1} \underline{z} - \overset{\text{lagrange multiplier}}{\mu} (\underline{z}^T \underline{z} - 1)$$



$\underline{x} = x\underline{i} + y\underline{j}$, $\underline{i} = \begin{bmatrix}1\\0\end{bmatrix}$, $\underline{j} = \begin{bmatrix}0\\1\end{bmatrix}$

$\underline{i}^T \underline{j} = 0$ ✓ , $\underline{i}^T \underline{i} = 1$ ✓

maximize $\Longrightarrow \dfrac{\partial F}{\partial \underline{Z}} = 2\,[\Sigma]_x^{-1}\,\underline{Z} - 2\mu\,\underline{Z} = \underline{0}$

$$\Rightarrow [\Sigma]^{-1}\,\underline{Z} = \mu\,\underline{Z}$$

$$\Rightarrow [\Sigma]\,\underline{Z} = \frac{1}{\mu}\,\underline{Z}$$

$\Rightarrow \boxed{[\Sigma]_x\,\underline{Z} = \lambda\,\underline{Z}} \Rightarrow$ بیشترین فاصله بیان مقدار ویژه
آن راستاست .

$\underline{\Phi}$ = eigen vector $(= \underline{Z})$ — بردار ویژه

λ = eigen value — مقدار ویژه

$\left|\,[\Sigma]_x - \lambda\,[I]\,\right| = 0 \longrightarrow \lambda = ?$ (مقدار ویژه) — معادله مشخصه

$\left([\Sigma]_x - \lambda\,[I]\right)\underline{\Phi} = 0 \longrightarrow \underline{\Phi} = ?$ (بردار ویژه)

ϕ_i = eigen vector corresponting to λ_i

$$\underline{\phi}_i^{T}\,\underline{\phi}_j^{T} = \begin{cases} K_i & i = j \\ 0 & i \neq j \end{cases} \Longleftrightarrow \text{Orthogonal}$$

$$[\Sigma]_x\,\underline{\Phi}_i = \lambda_i\,\underline{\Phi}_i \qquad \times\,\underline{\Phi}_j$$
$$[\Sigma]_x\,\underline{\Phi}_j = \lambda_j\,\underline{\Phi}_j \qquad \times\,\underline{\Phi}_i$$

$[\Sigma]$
قطری، متقارن $\dashrightarrow 0 = (\lambda_i - \lambda_j)\,\underline{\Phi}_i^{T}\,\underline{\Phi}_j$

$$\Rightarrow \begin{cases} i=j \longrightarrow 0=(\cancelto{0}{\lambda_i-\lambda_i})\,\underline{\phi}_i^T\,\underline{\phi}_i \xrightarrow{\text{طبق فرض}} \underline{\phi}_i^T\,\underline{\phi}_i=1 \\ i\neq j \longrightarrow 0=\overbrace{(\lambda_i-\lambda_j)}^{\neq 0}\,\underline{\phi}_i^T\,\underline{\phi}_j \longrightarrow \underline{\phi}_i^T\,\underline{\phi}_j=0 \end{cases}$$

x-Space •

Transformation Matrix $\equiv [a]$

$$[\Phi]\equiv[a]=\underbrace{\begin{bmatrix}\underline{\phi}_1 & \underline{\phi}_2 & \cdots & \underline{\phi}_n\end{bmatrix}}_{}$$

بردارهای ستونی / بردارهای ویژه ماتریس $[\Sigma]_x$

eigen vector matrix

$[\Phi]^T\cdot[\Phi]=[I]$

$[\Phi]^T=[\Phi]^{-1}$

orthogonal خواص

$$[\lambda]=\text{diag}\,\{\lambda_1\ \lambda_2\ \dots\ \lambda_n\}$$

$$=\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \emptyset \\ \emptyset & \ddots & \lambda_n\end{bmatrix}$$

eigen value matrix

$$\Rightarrow [\Sigma]_x\,\underline{z}=\lambda_i\,\underline{z}$$

$$\Rightarrow [\Sigma]_x\,[\Phi]=[\Phi]\,[\lambda]$$

• تبدیل برای قطری کردن ماتریس

$$\underline{Y}=[a]^T\underline{x}$$

$$[\Sigma]_y=[a]^T[\Sigma]_x[a]=[\Phi]^T[\Sigma]_x[\Phi]$$

$$[\Sigma]_x[\Phi]=[\Phi][\lambda] \xrightarrow{\times[\Phi]^T} [\Phi]^T[\Sigma]_x[\Phi]=[\lambda] \Big| \Rightarrow [\Sigma]_y=[\lambda]$$

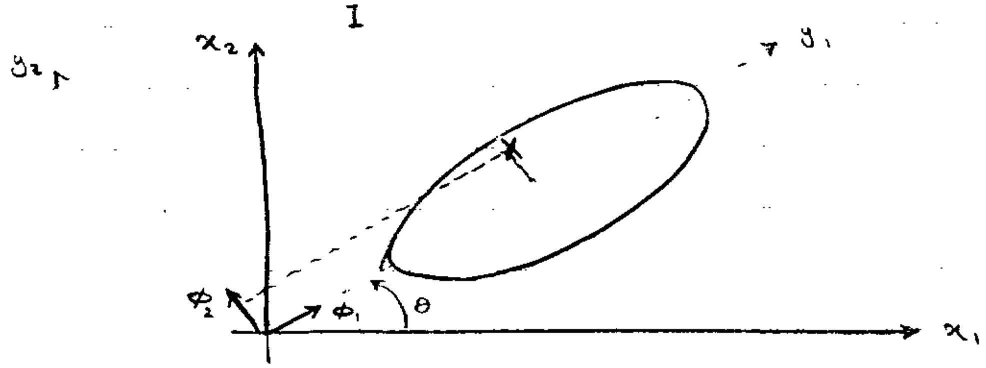
ماتریس قطری ← داده‌ها uncorrelated

برای قطری کردن ماتریس، ماتریس تبدیل [a] باید همان بردارهای ویژه ماتریس باشد ([Φ]).

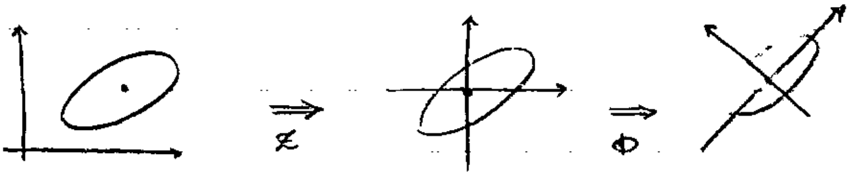
$$\underline{Y} = [\phi]^T \underline{X}$$

$$\underline{Y} = [\underline{\phi}_1 \ \underline{\phi}_2 \cdots \underline{\phi}_n]^T \underline{x}$$

$$Y_i = \underline{\phi}_i^T . \underline{X} = \underbrace{\|\phi_i\|}_{I} \ \|\underline{X}\| . \cos\theta = \|\underline{X}\| . \cos\theta$$



- coordinate transformation
- diagnolize $[\Sigma]$



درواقع

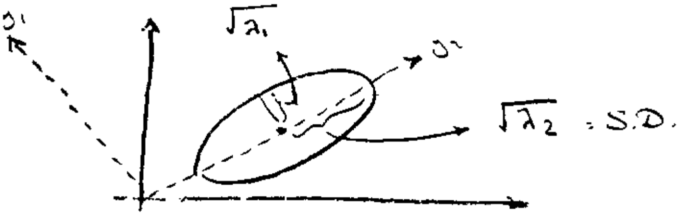
$$\underline{Y} = [\phi]^T \underline{x}$$

$$[\Sigma]_Y = [\lambda]$$

$$\|\underline{Y}\|^2 = \underline{Y}^T \underline{Y} = ([\phi]^T \underline{x})^T ([\phi]^T \underline{x}) = \underline{x}^T \overbrace{[\phi][\phi]^T}^{I} \underline{x} = \underline{X}^T \underline{X} = \|\underline{X}\|^2$$

In orthonormal transformation, distance is preserved.

$\lambda_i = var\{Y_i\}$



$[\Sigma]_y , [\lambda]$

$$d^2(y) = (\underline{y} - \underline{m}_y)^T [\Sigma]_y^{-1} (\underline{y} - \underline{m}_y)$$

$$d^2(y) = \sum_{i=1}^{n} \frac{(y_i - m_{iy})^2}{\lambda_i} \quad (*)$$

$$d^2(y) = \frac{(y_1 - m_{1y})^2}{\lambda_1} + \frac{(y_2 - m_{2y})^2}{\lambda_2} + \dots + \frac{(y_n - m_{ny})^2}{\lambda_n} = c$$

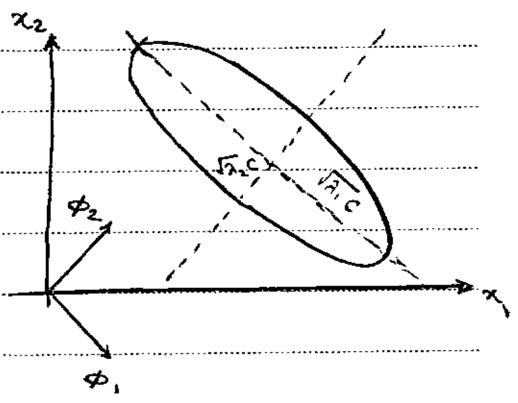
معادله بیضی ← نقطه‌ی بیضی معلوم است.

$$[\Sigma]_y^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & & 0 \\ & \frac{1}{\lambda_2} & \\ & & \ddots \\ 0 & & \frac{1}{\lambda_n} \end{bmatrix} \Longrightarrow (*)$$

$$P_y(y) = \frac{1}{(2\pi)^{n/2} |\Sigma|_y^{1/2}} \exp\{-\frac{1}{2} d^2(y)\} = k$$

☑

$$[\Sigma]_x = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \qquad \underline{m} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$



$$|[\Sigma] - \lambda [I]| = 0 \longrightarrow \lambda_1 = 3 , \lambda_2 = 1$$

$$\phi_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{+1}{\sqrt{2}} \end{bmatrix} \qquad \phi_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

$$([\Sigma] - 1[I])\phi_1 = (\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix})\phi_1 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \phi_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\phi_1 = \begin{bmatrix} a \\ b \end{bmatrix} \Rightarrow \begin{cases} a - b = 0 \\ -a + b = 0 \end{cases} \Rightarrow a = b \Longrightarrow$$

برای اینکه اندازه ϕ_1 برابر 1 باشد آن را نرمال می‌کنیم

$$\Rightarrow \phi_1 = \begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} \\ \frac{+b}{\sqrt{a^2+b^2}} \end{bmatrix} = \begin{bmatrix} \frac{a}{\sqrt{2a^2}} \\ \frac{b}{\sqrt{2b^2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Whitening Transformation *

برف :

$\underline{Y} = [a]^T \underline{X}$

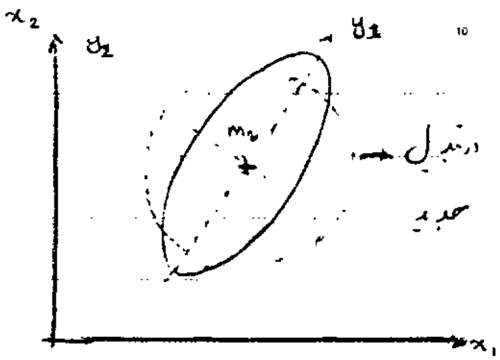
$[\Sigma]_y = [I]$

$[a] = [\phi][\lambda]^{-1/2}$

$[\phi]$ = eigen vector matrix of $[\Sigma]_x$

$[\lambda]$ = eigen value matrix of $[\Sigma]_x$

$$[\Sigma]_Y = [a]^T [\Sigma][a]$$
$$= ([\phi][\lambda]^{-1/2})^T [\Sigma]_x ([\phi][\lambda]^{-1/2})$$
$$= [\lambda]^{-1/2}[\phi]^T[\Sigma]_x[\phi][\lambda]^{-1/2}$$
$$= [\lambda]^{-1/2}[\lambda][\lambda]^{-1/2}$$
$$= [I]$$



مشخصات •

I. Whitening transformation is not orthonormal transformation

orthonormal $\Longleftrightarrow$ $[\underline{a}]^T[\underline{a}] = [\underline{I}]$

$$([\underline{\phi}][\underline{\lambda}]^{-1/2})^T([\underline{\phi}][\underline{\lambda}]^{-1/2}) = [\underline{\lambda}]^{-1/2}[\underline{\phi}]^T[\underline{\phi}][\underline{\lambda}]^{-1/2} = [\underline{\lambda}] \neq [I]$$

II. Euclidean distance is not preserved.

$$\|\underline{Y}\|^2 = \underline{y}^T\underline{y} = \underline{X}^T \underbrace{[\phi][\lambda]^{-1}[\phi]^T}\underline{X} = \underline{X}^T[\Sigma]_x\underline{X} \neq \|\underline{X}\|^2$$

III. After a whitening transformation, the covariance matrix is invariant under any orthonormal transformation.

$\underline{Y} = [\underline{a}]^T \underline{X}$

$[\Sigma]_Y = [I]$ → whitening transformation

$\underline{Z} = [\underline{a}]^T \underline{Y}$

$[\Sigma]_Z = [I]$ → orthonormal transformation

if $[\Psi] \equiv$ orthonormal transformation matrix

$$[\Psi]^T [\Sigma]_Y [\Psi] = [I]$$

Simultaneous Diagonalization • قطری کردن دو ماتریس

یعنی:

$[\Sigma]_1 \longrightarrow [I]$

$[\Sigma]_2 \longrightarrow [\lambda]$

1. find eigen value matrix and eigen vector matrix of $[\Sigma]_1$ then apply whitening transformation

$[\Phi] \equiv$ eigen vector of $[\Sigma]_1$

$[\mu] \equiv$ eigen value of $[\Sigma]_1$

$\underline{Y} = [\underline{a}]^T \underline{X}$

$[\underline{a}] = [\underline{\Phi}][\mu]^{-1/2}$

$[\mu]^{-1/2}[\Phi]^T[\Sigma]_1[\Phi][\mu]^{-1/2} = [I]$

$[\mu]^{-1/2}[\Phi]^T[\Sigma]_2[\Phi][\mu]^{-1/2} = [K]$

$[K]$ is not diagonal in general

2. find eigen value matrix & eigen vector matrix of $[K]$

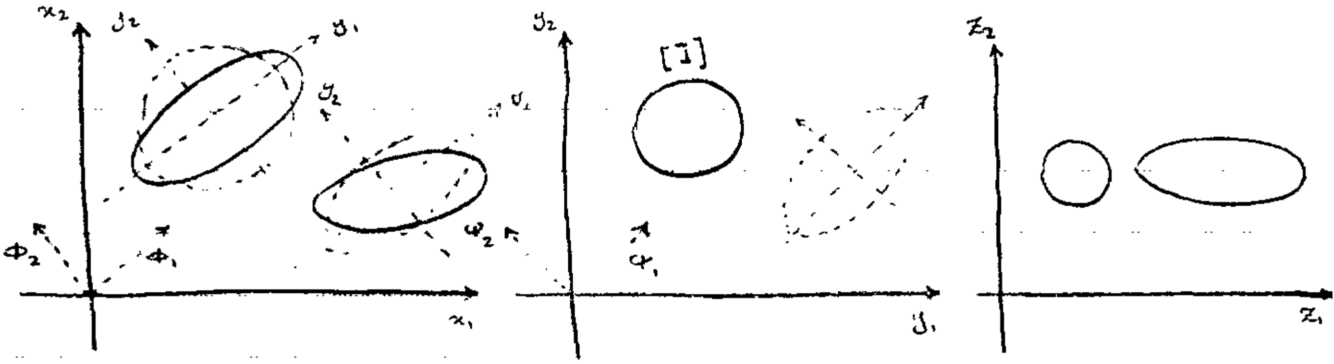
$\underline{Z} = [a]_2^T \underline{Y}$

$[a]_2^T \equiv [\Psi] \equiv$ eigen vector matrix of $[K]$

$[\lambda] \equiv$ eigen value matrix of $[K]$

$$[\Psi][\Sigma_Y]_1[\Psi] = [\Psi]^T[I][\Psi] = [I]$$

$$[\Psi][\Sigma_Y]_2[\Psi] = [\Psi]^T[K][\Psi] = [\lambda]$$



- تبدیل های قطری سازی همزمان

Theorem :

$$[\Sigma]_1 \longrightarrow [I]$$

$$[\Sigma]_2 \longrightarrow [\lambda] \equiv \text{diagonal } \{\lambda_1, \dots, \lambda_n\}$$

eigen vector matrix $([\Sigma]_1^{-1} \ [\Sigma]_2) \equiv [a] \rightarrow$ تبدیل های قطری سازی همزمان

eigen value matrix $([\Sigma]_1^{-1} \ [\Sigma]_2) \equiv [\lambda]$

پروژه‌های کامپیوتری : ۱ - ۲ - ۳ - ۴ - ۵ - ۶

مسائل : ۱ - ۲ - ۳ - ۴ - ۵ - ۶ - ۷ - ۱۰

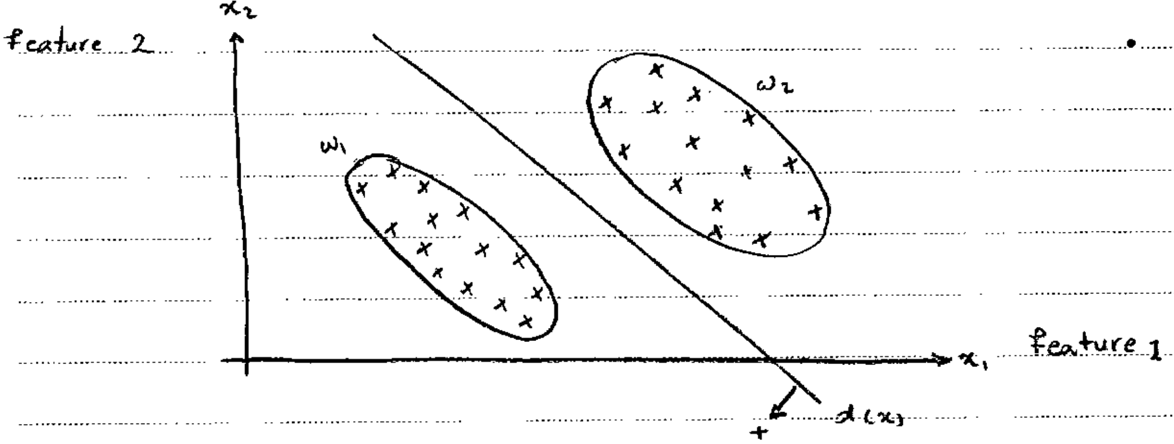
تحویل : ۲۳, ۷, ۸۷ ← ۱ هفته مهلت اضافی

Subject:
Year. Month. Date. 13

Decision Function

* توابع تصمیم گیری

(Discriminent Functions) توابع تمایز .



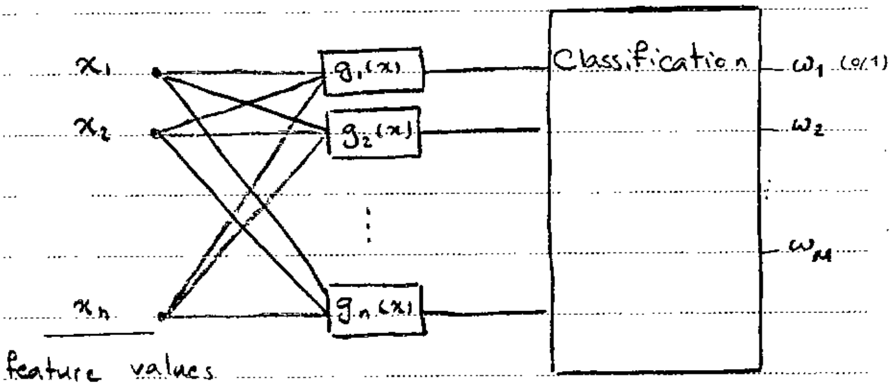
$d(x) = a_1 x_1 + a_2 x_2 + a_0$

$= W_1 x_1 + W_2 x_2 + W_0 = W^T x$

$$\left.\begin{array}{ll} d(x) > 0 & \text{belong } \omega_1 \\ d(x) < 0 & \text{belong } \omega_2 \end{array}\right\} \Rightarrow d(x) \overset{\omega_1}{\underset{\omega_2}{\gtrless}} 0$$

2-class problem

M - class problem



Indeterminant Region , IR محدوده .

Subject:

Year. Month. Date. ( )

*

I. M-Class Problem

$\underline{x} \in \omega_i$

$$d_i(\underline{x}) = \begin{cases} > 0 & \underline{x} \in \omega_i \\ < 0 & \text{o.w} \end{cases}$$

II. جداکردن کلاس i از کلاس j

$$\binom{M}{2} = \frac{M(M-1)}{2}$$

$\underline{x} \in \omega_i$

$$d_{ij}(\underline{x}) = \begin{cases} > 0 & \underline{x} \in \omega_i \\ < 0 & \underline{x} \in \omega_j \end{cases}$$

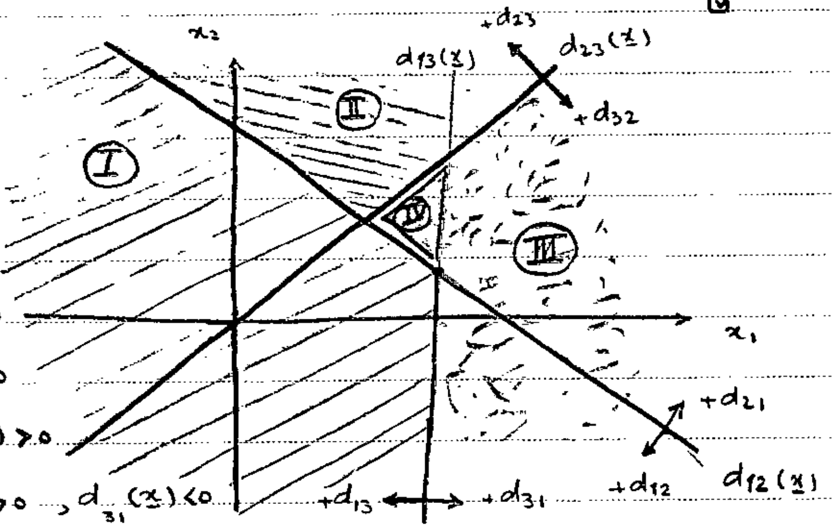
$d_{ij}(\underline{x}) = -d_{ji}(x)$

☑

$d_{12}(\underline{x}) = -x_1 - x_2 + 5$

$d_{13}(\underline{x}) = -x_1 + 3$

$d_{23}(\underline{x}) = -x_1 + x_2$



class I , $d_{12}(\underline{x}) > 0$, $d_{13}(\underline{x}) > 0$

class II : $d_{21}(\underline{x}) > 0$, $d_{23}(\underline{x}) > 0$

class III : $d_{31}(\underline{x}) > 0$, $d_{32}(\underline{x}) > 0$

class IV : IR $\rightarrow$ $\begin{cases} d_{32}(\underline{x}) > 0, \ d_{31}(\underline{x}) < 0 \\ d_{21}(\underline{x}) > 0, \ d_{23}(\underline{x}) < 0 \\ d_{13}(\underline{x}) > 0, \ d_{12}(\underline{x}) < 0 \end{cases}$

III. کلاس k ام

M Decision Function

i^{th} class : $d_i(\underline{x}) = \underline{w}_i^T \underline{x}$ → linear function

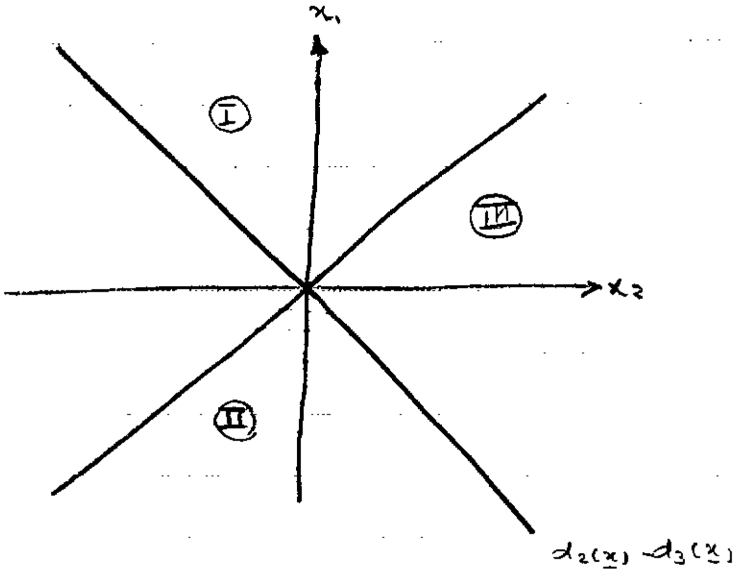
$\underline{x} \in \omega_i$ $d_i(\underline{x}) > d_j(\underline{x}) \quad \forall j \neq i$

لزومی ندارد تابع d عددی مثبت بدهد.

$$d_{ij}(\underline{x}) = d_i(\underline{x}) - d_j(\underline{x}) > 0$$
$$= w_i^T \underline{x} - w_j^T \underline{x} = (w_i - w_j)^T \underline{x} = \underline{w}_{ij}^T \underline{x}$$

$d_1(\underline{x}) = -x_1 + x_2$
$d_2(\underline{x}) = x_1 + x_2 - 1$
$d_3(\underline{x}) = -x_2$

$w_1 : \; d_1 > d_2 \; , d_1 > d_3$
$w_2 : \; d_2 > d_3 \; , d_2 > d_1$
$w_3 : \; d_3 > d_1 \; , d_3 > d_2$



حسن این روش اینست که IR وجود ندارد. مگر نقاطی که روی مرزها قرار دارند

Subject:
Year. Month. Date. 15

Classifiers *

Minimum Distance Classifier .

$\underline{x}$ unknown vector

$\underline{Z}_K$ = representative of ω_K

$D(\underline{x}, \omega_K) = D(\underline{x}, \underline{Z}_K)$

1 فاصله اقلیدسی $= \| \underline{x} - \underline{Z} \| = \sqrt{(x_1 - Z_1^k)^2 + \dots + (x_n - Z_n^k)^2}$

city block فاصله $= \sum |x_i - z_i| \longrightarrow$ مثلاً بین کاری ندارم

$$D^2(\underline{x}, \underline{Z}) = (x_1 - Z_1^K)^2 + \dots + (x_n - Z_n^K)^2$$
$$= \sum_{i=1}^{n} (x_i - Z_i^K)^2$$

$D^2(\underline{x}, Z_K)$ = Distance between $\underline{x}$ & ω_K

$$= \| \underline{x} - Z_K \|^2$$
$$= (\underline{x} - \underline{Z}_K)^T (\underline{x} - Z_K)$$
$$= \underline{x}^T \underline{x} - \underline{x}^T Z_K - Z_K^T \underline{x} + \underline{Z}_K^T \underline{Z}_K$$
$$= \underline{x}^T \underline{x} - 2 \underline{x}^T \underline{Z}_K + \underline{Z}_K^T Z_K$$

Classifier : $\underline{x} \in \omega_K$ if

$$D^2(\underline{x}, \underline{\omega}_K) = \min_{\forall i} \left[\underbrace{\underline{x}^T \underline{x}} - 2 \underline{x}^T \underline{Z}_i + Z_i^T Z_i \right]$$

این جمله همیشه ثابت است، پس آنرا حذف می‌کنیم و حد است نه تابع خواهد بود.

$$D^2(\underline{x}, \underline{\omega}_K) = \min \left[-2 \underline{x}^T Z_i + \underline{Z}_i^T \underline{Z}_i \right]$$
$$= \max \left[2 \underline{x}^T \underline{Z}_i - \underline{Z}_i^T \underline{Z}_i \right]$$

$$d_K(\underline{x}) = \underline{x}^T Z_K - \frac{1}{2} Z_k^T Z_k$$

Subject:
Year. Month. Date. ( )

$$\underline{x} \in \omega_k \qquad d_k(\underline{x}) > d_j(\underline{x}) \quad \forall j \neq k$$

$$d_{kj}(\underline{x}) = d_k(\underline{x}) - d_j(\underline{x})$$

یادآوری : تابع خطی در فضای n بعدی

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \Rightarrow \underline{x}' = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\underline{w} = \begin{bmatrix} w_1 \\ \cdot \\ w_n \end{bmatrix} \Rightarrow \underline{w}' = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$d(\underline{x}) = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + w_0$$
$$= \sum_{i=1}^{n} w_i x_i + w_0$$
$$= \underline{w}^T \underline{x} + w_0 = \underline{w}'^T \underline{x}'$$

پس :

$$d_k(\underline{x}) = \underline{x}^T \underline{z}_k - \frac{1}{2} z_k^T z_k$$

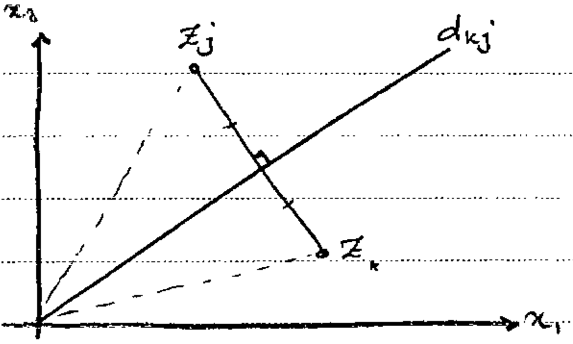
$$d(\underline{x}) = \underset{\text{weight}}{\underline{w}^T} \underline{x} \qquad \begin{array}{l} \longrightarrow \underline{w} = \underline{z} \\ \searrow w_0 = -\frac{1}{2} z_k^T z_k \end{array}$$

$$d_k(\underline{x}) = \underline{w}_k^T \underline{x} = \underline{x}^T \underline{z}_k - \frac{1}{2} z_k^T z_k$$

$$d_{kj}(\underline{x}) = d_k(\underline{x}) - d_j(\underline{x})$$ جداکننده کلاس k و j

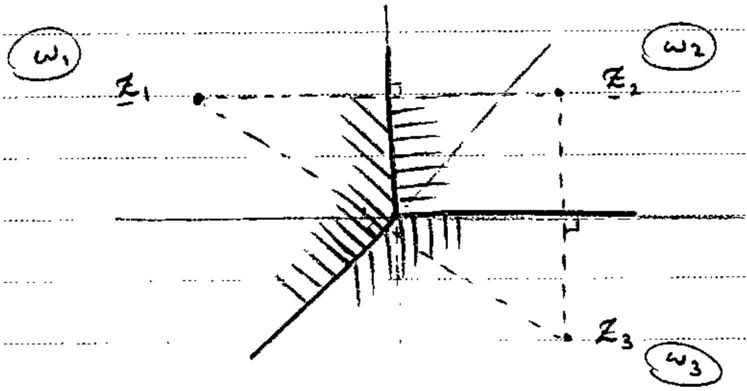
$$= \underline{x}^T (\underline{z}_k - \underline{z}_j) - \frac{1}{2} (z_k^T z_k - z_j^T \underline{z}_j)$$

$$= \underline{x}^T (\underline{z}_k - \underline{z}_j) - \frac{1}{2} (z_k - z_j)^T (\underline{z}_k + \underline{z}_j)$$



تابع تصمیم گیری عمود منصف پاره‌خطی است که دو نقطه Z_k و Z_j را وصل می‌کند.

Z_i نماینده یک کلاس است و معمولاً میانگین کلاس‌هاست.



اگر داده‌ها uni-Model باشند، این classifier بهترین انتخاب است.

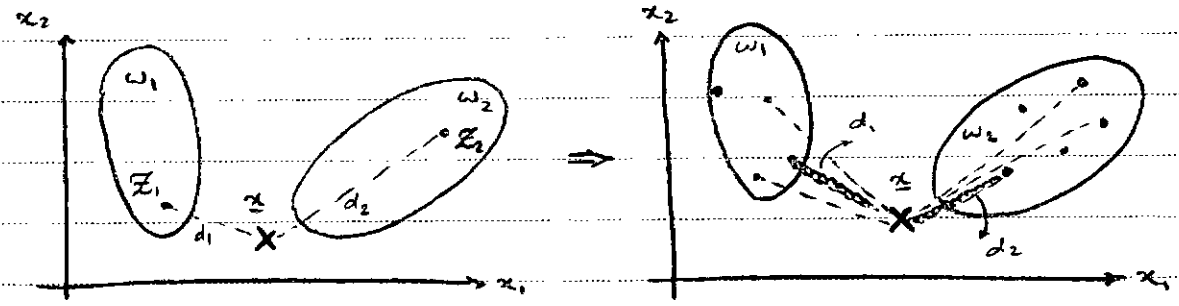
اگر بجای یک نماینده، M نماینده از هر کلاس موجود باشد:

$$D(\underline{x}, \omega_k) = \min_{m=1,\dots,N_k} \left[D(\underline{x}, Z_k^{(m)}) \right]$$

k^{th} class is represented by N_k representatives (prototypes)

$$d_k(\underline{x}) = \max_{m=1,\dots,N_k} \left[\underline{x}^T Z_k^{(m)} - \frac{1}{2} \left| Z_k^{(m)} \right|^2 \right]$$

$k = 1, \dots, M \sim$ # of classes



$\underline{x} \in Z_1$ $\underline{x} \in Z_1$

آیا داده‌ها به طور خطی جداپذیر هستند ؟

☑ اگر داده‌ها یک تجمع convex باشند ✓

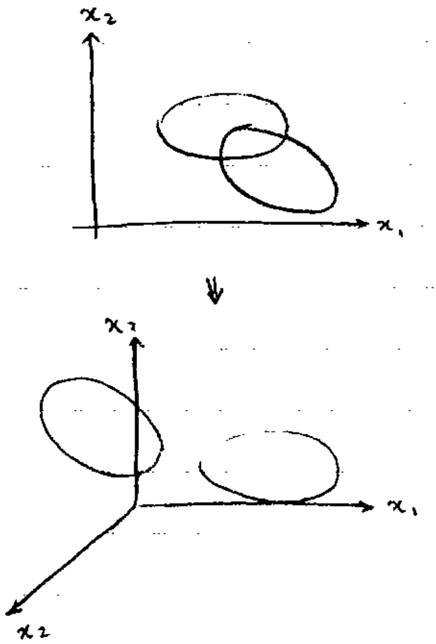
... ← نقطه‌ای زیادی وجود دارند

☑ اگر داده‌ها دارای توزیع نرمال باشند ✓

- Generalized Dicision Function

$$d(\underline{x}) = W_1 f_1(\underline{x}) + W_2 f_2(\underline{x}) + \dots + W_k f_k(\underline{x})$$

$$= \sum_{i=1}^{k} W_i f_i(\underline{x})$$



☑ 2-D

$$d(\underline{x}) = W_{11} x_1^2 + W_{22} x_2^2 + W_{12} x_1 x_2 + W_{21} x_2 x_1 + W_1 x_1 + W_2 x_2 + W_0$$

$$\left.\begin{array}{l} x_1^2 \longrightarrow f_1(\underline{x}) \\ x_2^2 \longrightarrow f_2(\underline{x}) \\ x_1 x_2 \longrightarrow f_3(\underline{x}) \\ \vdots \end{array}\right| \Rightarrow \quad \underline{x}^* = \begin{bmatrix} f_1(\underline{x}) \\ f_2(\underline{x}) \\ \vdots \\ f_k(\underline{x}) \\ 1 \end{bmatrix} \qquad \omega_0 \swarrow$$

☑ n-D : quadratic

$$d(\underline{x}) = \underbrace{\sum_{j=1}^{n} W_{jj} x_j^2}_{n} + \underbrace{\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} W_{jk} x_j x_k}_{\frac{n(n-1)}{2}} + \underbrace{\sum_{j=1}^{n} W_j x_j}_{n} + \underbrace{W_{n+1}}_{1}$$

$\frac{(n+1)(n+2)}{2}$ = تعداد ضرایب ←

$$N_w = \frac{(n+r)!}{r!\, n!}$$: r-order , n-D حالت کلی :

$n=2$ $r=2$ $N_w=6$ ☑

$n=3$ $r=2$ $N_w=10$

$n=3$ $r=4$ $N_w=35$

$$d(\underline{x}) = \underline{x}^T [A] \underline{x} + \underline{x}^T \underline{b} + c$$

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$[A] = (a_{jk}) \Longrightarrow \left| \begin{array}{lll} w_{jj} = a_{jj} & j=1, \dots, n & \\ w_{jk} = 2\, a_{jk} & j=k, \dots, n & i \neq k \\ w_j = b_j & & \\ w_{n+1} = c & & \end{array} \right.$$

- $\star$ $[A]=[I]$ $\Longrightarrow$ hyper sphere
- $\star$ $[A] > 0$ $\Longrightarrow$ hyper elepsoid
- $\star$ $[A] \geqslant 0$ $\Longrightarrow$ hyper elepsoid cylenderical

Statistical Discriminant Function — * توابع تمایز آماری :

Hypothesis Test for 2-I class •

Bayes Eq. → احتمال ثانوی

$$P(\omega_1 \mid \underline{x}) \equiv q_1(\underline{x}) = P_1 \frac{P_1(\underline{x})}{P(x)}$$

$P_i \equiv P(\omega_i) = \text{Prob}\{\omega_i\} \quad i = 1, 2$ → احتمال اولیه apriori prob. of wi)

$P_i(x) = P(\underline{x} \mid \omega_i)$ → تابع چگالی احتمال شرطی برای کلاس wi

$P(\underline{x}) = \sum_{i=1}^{M=2} P_i . P_i(\underline{x})$ → تابع چگالی احتمال کلی

Bayes Classifier:

$\underline{x}$ unknown

$$q_1(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} q_2(\underline{x}) \qquad .I.$$

$$\frac{P_1 . P_1(\underline{x})}{P(\underline{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P_2 . P_2(\underline{x})}{P(\underline{x})} \Longrightarrow P_1 . P_1(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P_2 . P_2(\underline{x}) \qquad .II.$$

$\ell(\underline{x})$. liklihood function $= \frac{P_1(x)}{P_2(x)} \Longrightarrow \ell(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P_2}{P_1}$ …… threshold .III.

$h(\underline{x}) = -\ln \ell(\underline{x}) = \ln P_2(x) - \ln P_1(x) \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \ln\left(\frac{P_1}{P_2}\right)$ .IV.

↳ statistical discriminant function

$\underline{X}$ with binary valued elements & independent ✓

$\underline{x} = [\, x_1 \;\ldots\; x_n \,]^T$

2 - class problem

$\text{Prob}\{ x_i = 1 \mid \omega_1 \} = P_i'$

$\text{Prob}\{ x_i = 0 \mid \omega_1 \} = 1 - P_i'$

$\text{Prob}\{ x_i = 1 \mid \omega_2 \} = q_i'$

$\text{Prob}\{ x_i = 0 \mid \omega_2 \} = 1 - q_i'$

$P(\underline{x} \mid \omega_i) = P(x_1 \mid \omega_i) \cdot P(x_2 \mid \omega_i) \cdot \ldots \cdot P(x_n \mid \omega_i)$

$P(\underline{x} \mid \omega_1) = \prod_{i=1}^{n} (P_i')^{x_i} (1 - P_i')^{1-x_i}$

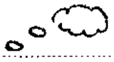
$P(\underline{x} \mid \omega_2) = \prod_{i=1}^{n} (q_i')^{x_i} (1 - q_i')^{1-x_i}$

$$\ell(\underline{x}) = \frac{P_1(\underline{x})}{P_2(\underline{x})} = \frac{\prod (P_i')^{x_i}(1-P_i')^{1-x_i}}{\prod (q_i')^{x_i}(1-q_i')^{1-x_i}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P_2}{P_1}$$

$$\ln \ell(\underline{x}) = \ln \frac{P_1(\underline{x})}{P_2(\underline{x})} = \sum \left[x_i \ln\left(\frac{P_i'}{q_i'}\right) + (1-x_i) \ln \frac{1-P_i'}{1-q_i'} \right] + \ln \frac{P_1}{P_2} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$$h(\underline{x}) = \sum_{i=1}^{n} \left[x_i \ln \frac{P_i'(1-q_i')}{q_i'(1-P_i')} + \ln \frac{1-P_i'}{1-q_i'} \right] + \ln \frac{P_1}{P_2}$$

آبیاری



$$\underline{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \xrightarrow{d(\underline{x}) = \underline{w}^T \underline{x}} \underline{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$ ⌒ agumented weight vector

$$h(\underline{x}) = \underline{w}^T \underline{x} = \sum_{i=1}^{n} x_i \cdot w_i + w_0 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$\underline{X}$ with 3-valued elements & independent

$x_i = 1 , 0 , -1$

$h(\underline{x}) = ?$

2-class problem

$\underline{X} = \begin{bmatrix} x \\ y \end{bmatrix}$ $\quad \underline{m}^i_x = \begin{bmatrix} m_x \\ m_y \end{bmatrix}$ $\quad [\Sigma]^i_x = \begin{bmatrix} \sigma_x^2 & \rho_{xy} \cdot \sigma_x \cdot \sigma_y \\ \rho_{xy} \cdot \sigma_y \cdot \sigma_x & \sigma_y^2 \end{bmatrix}$

$\underline{X} = N(\underline{m} , [\Sigma])$

$|\Sigma| = \sigma_x^2 \sigma_y^2 (1 - \rho_{xy}^2)$

$$p(x,y \mid \omega_i) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho_{xy}^2}} \cdot \exp\left\{ \frac{-1}{2(1-\rho_{xy}^2)} \left[\left(\frac{x-m_x}{\sigma_x}\right)^2 + \left(\frac{y-m_y}{\sigma_y}\right)^2 - \frac{2\rho_{xy}(x-m_x)(y-m_y)}{\sigma_x \sigma_y} \right] \right\}$$

$d(\underline{x}) = (\underline{x} - \underline{m}_x)^T [\Sigma]_x^{-1} (\underline{x} - m_x)$ ↗

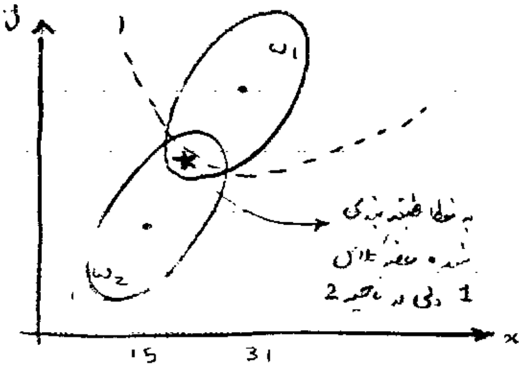
$P_1 = P(\omega_1) = 0.8$ $\quad \underline{m}_1 = \begin{bmatrix} 26 \\ 85 \end{bmatrix}$ $\quad m_2 = \begin{bmatrix} 22 \\ 70 \end{bmatrix}$ $\quad \rho^1_{xy} = 0.6$

$P_2 = P(\omega_2) = 0.2$ $\quad \rho^2_{xy} = 0.5$

$[\Sigma]_1 = \mathrm{cov}\{\underline{X} \mid \omega_1\} = \begin{bmatrix} 4 & 6 \\ 6 & 25 \end{bmatrix}$ $\quad [\Sigma]_2 = \begin{bmatrix} 9 & 12 \\ 12 & 64 \end{bmatrix}$

$h(\underline{x}) = -5.819 x^2 + 3.167 xy$
$- y^2 + 41.09 \; x$
$+ 27.30 \; y - 5746$

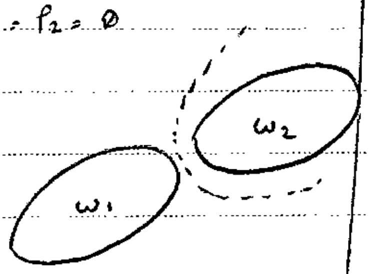
[illegible]



$\sigma_{1x} = \sigma_{2x}$

$\sigma_{1y} > \sigma_{2y}$

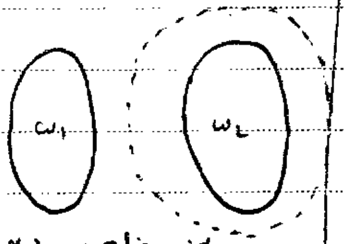
$\rho_1 = \rho_2 = 0$



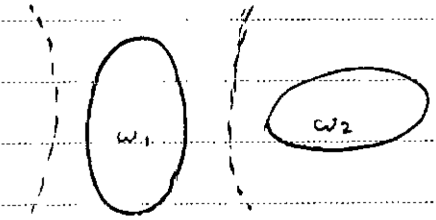
$h(\underline{x}) \sim$ parabolic

$\frac{\sigma_{1x}}{\sigma_{2x}} = \frac{\sigma_{1y}}{\sigma_{2y}}$

$\rho_1 = \rho_2 = 0$



$h(\underline{x}) \sim$ elipsoid

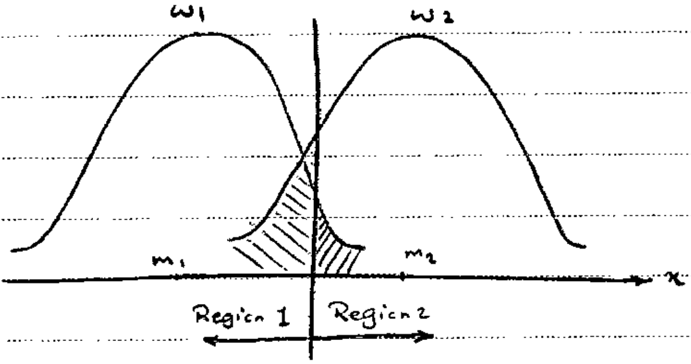


$h(\underline{x}) \sim$ hyperbolic

- خطای طبقه بندی

1-D

2-class problem



انواع خطا: عضو کلاس 1 در ناحیه 2

عضو کلاس 2 در ناحیه 1

ε: Probability of Error

$$P_1 . p_1(x) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P_2 . p_2(x)$$

انتگرال n-تایی

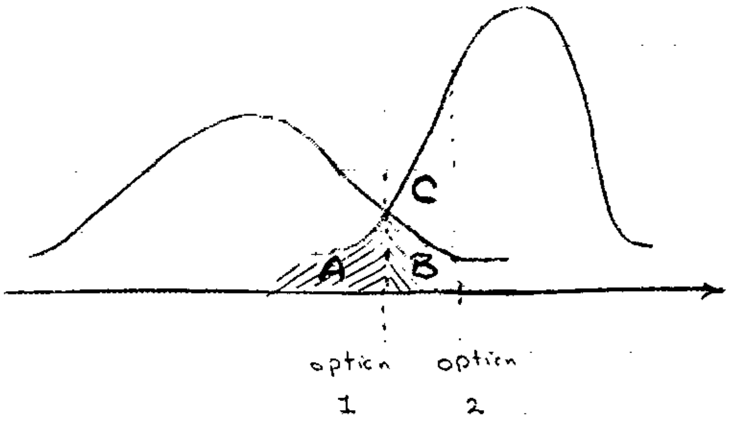
$$\varepsilon = \int_{R_2} P_1 . p_1(\underline{x}) . d\underline{x} + \int_{R_1} P_2 . p_2(\underline{x}) . d\underline{x}$$

$$= P_1 \underbrace{\int_{R_2} p_1(\underline{x}) . d\underline{x}} + P_2 \underbrace{\int_{R_1} p_2(\underline{x}) . d\underline{x}}$$

$$= P_1 \varepsilon_1 + P_2 \varepsilon_2$$

$$h(\underline{x}) = -\ln \ell(x) \overset{\omega_2}{\underset{\omega_1}{\gtrless}} 0$$

توابع Bayes ، بهترین نوع تابع مرز تصمیم گیری هستند
این تابع بهینه است، چون:



option 1 : $\varepsilon = A + B$

option 2 : $\varepsilon = A + B + C$

option 1 ✓ ⟹ Bayes ✓

✎ این قضیه را با ریاضی اثبات نمایید.

$$h(\underline{x}) = -\ln \ell(\underline{x}) = \ln P_2(\underline{x}) - \ln P_1(\underline{x}) \overset{\omega_2}{\underset{\omega_1}{\gtrless}} \ln \frac{P_1}{P_2}$$

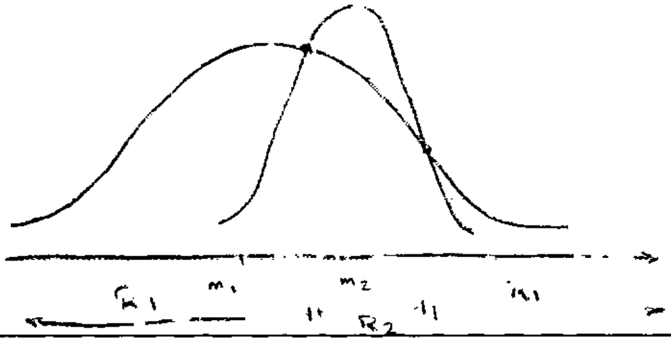
$h(\underline{x}) \longrightarrow$ Random Variable

با محاسبه خطای $h(\underline{x})$ می‌توان خطای چندبعدی $l(\underline{x})$ را یک بعدی بدست آورد

$$\varepsilon = P_1 \int_{R_2} P_H(h|\omega_1) \cdot dh + P_2 \cdot \int_{R_1} P_H(h|\omega_2) \cdot dh$$

↪ انتگرال یک بعدی

قسمت دشوار این مسئله، تعیین تابع $h(\underline{x})$ است.



$$l(\underline{x}) = \frac{P_1(\underline{x})}{P_2(\underline{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P_2}{P_1}$$

$$h(\underline{x}) = -\ln l(x) = \ln P_2(\underline{x}) - \ln P_1(x) \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \ln \frac{P_1}{P_2}$$

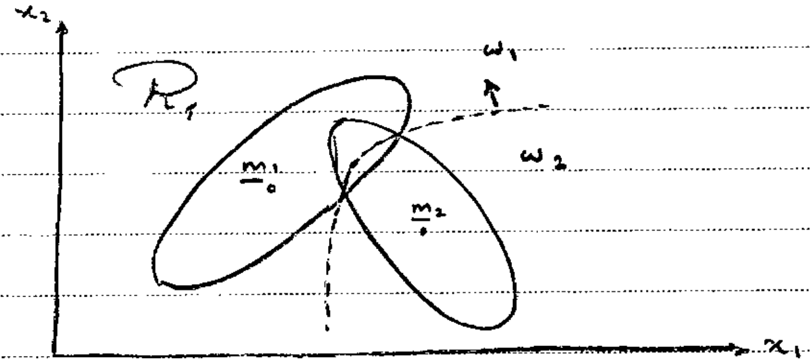
$\omega_1 : N(\underline{m}_1, [\Sigma]_1)$ $\qquad P_1$ ☑

$\omega_2 : N(\underline{m}_2, [\Sigma]_2)$ $\qquad P_2$

n - D , 2-class problem

$$P_i(\underline{x}) = \frac{1}{(2\pi)^{n/2} |[\Sigma]_i|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{m}_i)^T [\Sigma]_i^{-1} (\underline{x}-\underline{m}_i)\right\} \qquad i=1,2$$

$$\begin{aligned} h(\underline{x}) = -\ln l(\underline{x}) = & \frac{1}{2}(\underline{x}-\underline{m}_1)^T [\Sigma]_1^{-1} (\underline{x}-\underline{m}_1) \\ & - \frac{1}{2}(\underline{x}-\underline{m}_2)^T [\Sigma]_2^{-1} (\underline{x}-\underline{m}_2) \\ & + \frac{1}{2} \ln \frac{|[\Sigma]_1|}{|[\Sigma]_2|} \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \ln \frac{P_1}{P_2} \end{aligned}$$

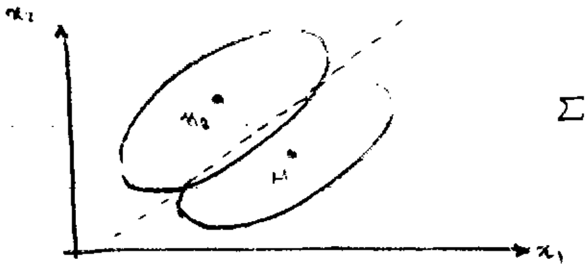


فرم تابع نرمال با
کوواریانس های متفاوت
≡
شکل تابع جداکننده
درجه 2 است.

* Si $[\Sigma]_1 = [\Sigma]_2 = [\Sigma]$

$$h(\underline{x}) = (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} \underline{x} + \frac{1}{2}(\underline{m}_1^T [\Sigma]^{-1} \underline{m}_1 - m_2^T [\Sigma]^{-1} m_2) \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \ln \frac{P_1}{P_2}$$

$h(\underline{x}) = \underline{W}^T \underline{x} + \omega_0$ → تابع خطی



$$\Sigma_1 = \Sigma_2$$

Special Case ☑

$\underline{m}_i = 0 \qquad i = 1, 2$

$$[\Sigma]_i = \begin{bmatrix} 1 & \rho_i & \rho_i^2 & \cdots & \rho_i^{n-1} \\ \rho_i & 1 & & & \\ \rho_i^2 & & 1 & & \\ & & & \ddots & \rho_i \\ \rho_i^{n-1} & & & & 1 \end{bmatrix}$$

→ toeplitz matrix → 1 قطراصلی

→ ρ_i : correlation coefficient of random variables for i^{th} class

$c_{ij} = \rho_{ij} \cdot \sigma_i \cdot \sigma_j$

خواص ماتریس toeplitz →

$$[\Sigma]_i^{-1} = \frac{1}{1-\rho_i^2} \begin{bmatrix} 1 & -\rho_i & 0 & \cdots & 0 \\ -\rho_i & 1+\rho_i^2 & -\rho_i & & 0 \\ 0 & -\rho_i & 1+\rho_i^2 & & \vdots \\ \vdots & & & \ddots & -\rho_i \\ 0 & 0 & \cdots & -\rho_i & 1+\rho_i^2 \end{bmatrix}$$

$$|\Sigma_i| = (1-\rho_i^2)^{n-1}$$

$$h(\underline{x}) = -\ln \ell(\underline{x}) = \left(\frac{1+\rho_1^2}{1-\rho_1^2} - \frac{1+\rho_2^2}{1-\rho_2^2}\right) \sum_{i=1}^{n} x_i^2 - \left(\frac{\rho_1^2}{1-\rho_1^2} - \frac{\rho_2^2}{1-\rho_2^2}\right)(x_1^2 + x_n^2)$$

$$-\left(\frac{2\rho_1}{1-\rho_1^2} - \frac{2\rho_2}{1-\rho_2^2}\right) \sum_{i=1}^{n-1} x_i x_{i+1}$$

$$+(n-1)\ln\frac{1-\rho_1^2}{1-\rho_2^2} \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \ln\frac{P_1}{P_2}$$

نقش sample اول و آخر در واقع [illegible] با بقیه [illegible]
بررسی شده

Subject:
Year. Month. Date. 21

$$h(\underline{x}) = \frac{\sum_{i=1}^{n} x_i \cdot x_{i+1}}{\sum_{i=1}^{n} x_i^2} \gtrless t$$

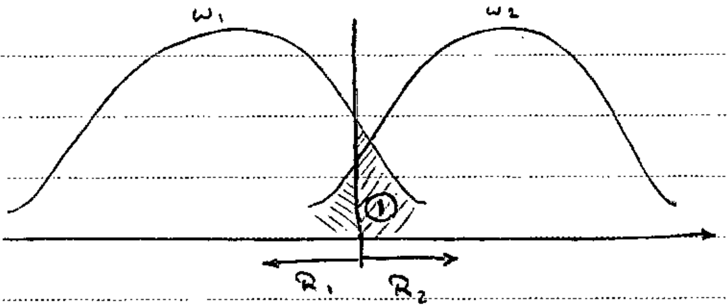
عامل تعیین کننده تمایز ، ماتریس covarience

* حالت دارای هزینه در طبقه‌بندی

C_{12} = cost of deciding $\underline{x} \in \omega_2$ when $\underline{x} \in \omega_1$ → wrong decision
C_{11} = cost of deciding $\underline{x} \in \omega_1$ when $\underline{x} \in \omega_1$
C_{21} = cost of deciding $\underline{x} \in \omega_1$ when $\underline{x} \in \omega_2$ → wrong decision
C_{22} = cost of deciding $\underline{x} \in \omega_2$ when $\underline{x} \in \omega_2$

$C_{ij} \longrightarrow$ i: true class j: decision

$C_{12} > C_{11}$ و $C_{21} > C_{22}$ ⟶ خطا هزینه بر است



$r = E\{\text{cost}\} \longrightarrow$ To minimize ...

$$r = \int_{R_1} C_{11} P_1 \cdot p_1(\underline{x})\, d\underline{x} + \int_{R_2} C_{12} \cdot P_1 \cdot p_1(x)\, d\underline{x} + \int_{R_1} C_{21} P_2 \cdot p_2(\underline{x})\, d\underline{x} + \int_{R_2} C_{22} P_2 p_2(\underline{x})\, d\underline{x}$$

$$= \int_{R_1} \{C_{11} P_1 p_1(\underline{x}) + C_{21} P_2 p_2(\underline{x})\}\, d\underline{x} + \underbrace{\int_{R_2} \{C_{12} P_1 p_1(\underline{x}) + C_{22} P_2 p_2(\underline{x})\}\, dx}_{②}$$

$$① = \int_{R_2} p_1(\underline{x})\, d\underline{x} = 1 - \int_{R_1} p_1(\underline{x})\, d\underline{x}$$

$$ⓡ = C_{12} P_1 \left[1 - \int_{R_1} p_1(\underline{x})\, d\underline{x}\right] + C_{22} P_2 \left[1 - \int_{R_1} p_2(\underline{x}) . d\underline{x}\right]$$

$$r = \underbrace{C_{12} P_1}_{cte} + \underbrace{C_{22} P_2}_{cte} + \int_{R_1} \left\{ -\underbrace{(C_{12} - C_{11})}_{>0} P_1 p_1(\underline{x}) + \underbrace{(C_{21} - C_{22})}_{>0} P_2 p_2(\underline{x}) \right\} . d\underline{x}$$

$$g(\underline{x}) > 0$$

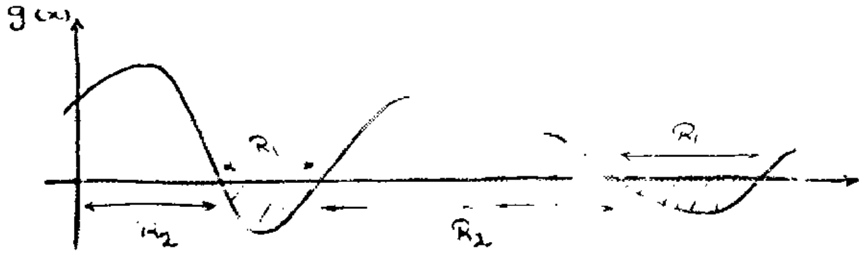
$r \rightarrow$ to be minimized $\Rightarrow$ $g \rightarrow$ to be minimized

$$r = \arg \min \left\{ \int_{R_1} g(\underline{x}) . d\underline{x} \right\}$$

$$\exists \underline{x} \Rightarrow g(\underline{x}) < 0 \longrightarrow \underline{x} \in \omega_1 \quad ✓$$

$$(C_{12} - C_{11}) P_1 \, p_1(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} (C_{21} - C_{22}) P_2 \, p_2(\underline{x})$$

$$\ell(\underline{x}) = \frac{p_1(\underline{x})}{p_2(\underline{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P_2 (C_{21} - C_{22})}{P_1 (C_{12} - C_{11})}$$ - threshold



Subject:
Year. Month. Date. 22

1-D 2-class Problem ☑

$P(x|\omega_1) = \frac{1}{\sqrt{2\pi}\sqrt{3}} e^{-1/2(\frac{x^2}{3})}$ $E\{X|\omega_1\} = 0$ $var\{X|\omega_1\} = 3$

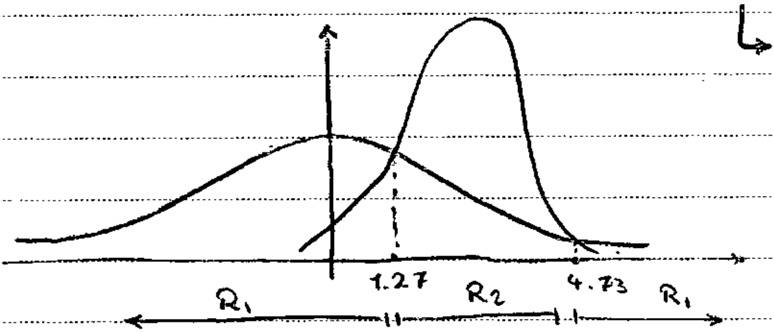
$P(x|\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}$ $E\{X|\omega_2\} = 2$ $var\{x|\omega_2\} = 1$

$P_1 = P_2 = \frac{1}{2}$ $C_{11} = C_{22} = 0$ $C_{21} = 1$ $C_{12} = \sqrt{3}$

$$\ell(\underline{x}) = \frac{P_1(\underline{x})}{P_2(\underline{x})} = \frac{e^{-\frac{1}{2}\frac{x^2}{3}}}{e^{-\frac{1}{2}(x-2)^2}} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 1$$

$$\xrightarrow{Ln} -\frac{1}{2}\frac{x^2}{3} + \frac{1}{2}(x-2)^2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0 \longrightarrow 2x^2 - 12x + 12 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$\hookrightarrow x_1 = 4.73, \; x_2 = 1.27$



The Neyman - Pearson Test *

• انواع خطا در تصمیم گیری

$\varepsilon_1 = \int_{R_2} P_1(\underline{x}) . d\underline{x}$

$\varepsilon_2 = \int_{R_1} P_2(\underline{x}) . d\underline{x}$

• مسأله

حل: $\varepsilon_2 = \varepsilon_0 \equiv cte \longrightarrow \varepsilon_1 \equiv$ minimize

Minimize ε_1 subject to $\varepsilon_2 = \varepsilon_0$

تابع معیاری: $r = \varepsilon_1 + \mu(\varepsilon_2 - \varepsilon_0)$

$$r = \varepsilon_1 + \mu\,(\varepsilon_2 - \varepsilon_0)$$

$$= \int_{R_2} P_1(\underline{x})\,d\underline{x} + \mu\left\{\left(\int_{R_1} P_2(\underline{x})\,d\underline{x}\right) - \varepsilon_0\right\}$$

$$= \underbrace{(1-\mu\,\varepsilon_0)}_{cte} + \int_{R_1}\left(\mu\,P_2(\underline{x}) - P_1(\underline{x})\right)d\underline{x}$$

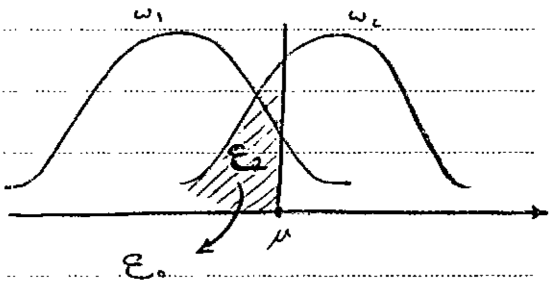
تابع مؤثر در کمینه کردن $\rightarrow\ <0 \Rightarrow \omega_1$

$$P_1(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \mu\,P_2(\underline{x}) \longrightarrow \boxed{\ell(\underline{x}) = \frac{P_1(\underline{x})}{P_2(\underline{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \mu}$$

• محاسبه: $\mu \leftarrow$ Lagrange Multiplier

$$\varepsilon_2 = \int_{R_1} P_2(\underline{x})\,d\underline{x} = \varepsilon_0 \rightarrow cte$$

$$\varepsilon_2 = \int_{-\infty}^{\mu} P_H(h \mid \omega_2)\,dh = \varepsilon_0$$



$\Leftarrow$ با تغییر μ می‌توان ε_0 را بدست آورد.

2-D ☑

2 - Normally Distributed Class

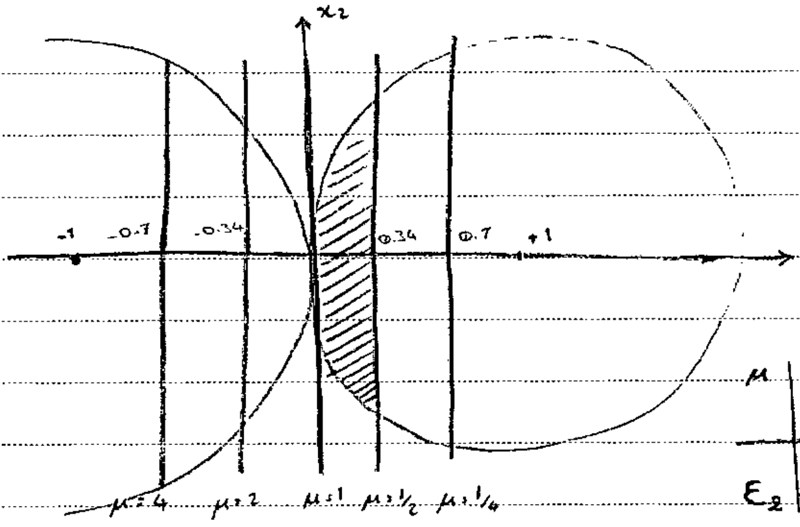
$$\underline{m}_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \qquad \underline{m}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad [\Sigma]_1 = [\Sigma]_2 = [I]$$

$$P_1 = P_2 = \frac{1}{2}$$

$$[\Sigma_1] = [\Sigma]_2 \Longrightarrow h(\underline{x}) = (\underline{m}_2 - \underline{m}_1)^T[\Sigma]^{-1}\underline{x} + \frac{1}{2}\left(\underline{m}_1^T[\Sigma]^{-1}\underline{m}_1 - \underline{m}_2^T[\Sigma]^{-1}\underline{m}_2\right) \gtrless \ln\frac{P_1}{P_2}$$

$$\ell(\underline{x}) = \frac{P_1(\underline{x})}{P_2(\underline{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \mu \Rightarrow h(\underline{x}) = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right)^T \underset{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}{\underline{x}} + \frac{1}{2}\left([-1\ 0]\begin{bmatrix} -1 \\ 0 \end{bmatrix} - [1\ 0]\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right. \gtrless -\ln\mu$$

$$h(\underline{x}) = 2x_1 \underset{\omega_1}{\overset{\omega_2}{\gtrless}} -\ln\mu$$



$$\varepsilon_2 = \int_{R_1} P_2(\underline{x})\,d\underline{x} = \varepsilon_0$$

μ	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$
ε_2	0.04	0.09	0.16	0.25	0.38

Error Table 3.1 براساس

The Minmax Test *

- صورت مسئله

$$\ell(\underline{x}) = \frac{P_1(\underline{x})}{P_2(\underline{x})} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P_2}{P_1}$$

اگر احتمال اولیه (P_i) تغییر کند، تصمیم گیری چگونه باید باشد.
☑ در مثال ماهی ها، با تغییر احتمال نوع ماهی در منطقه ی مختلف

-

$$r = E\{cost\}$$

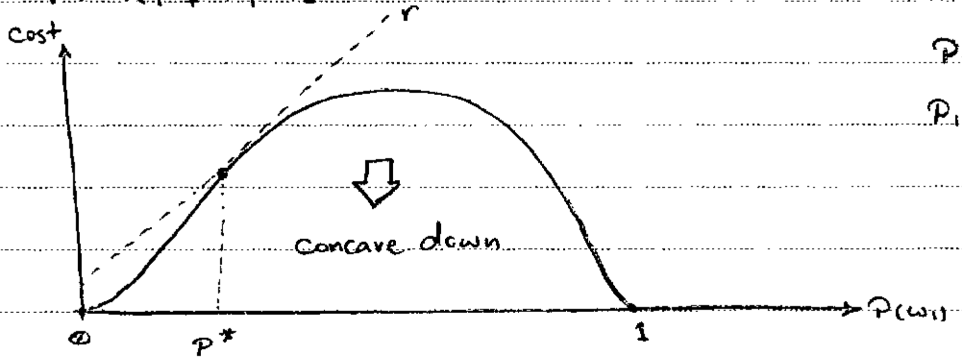
$$r = \int_{R_1} \{C_{11} P_1\, p_1(\underline{x})\,d\underline{x} + C_{21} P_2\, p_2(\underline{x})\,d\underline{x}\}\,d\underline{x} + \int_{R_2} \{C_{22} P_2\, p_2(\underline{x}) + C_{12} P_1\, p_1(\underline{x})\}\,d\underline{x}$$

$$r = C_{12} P_1 + C_{22} P_2 + \int_{R_1} \{-(C_{12}-C_{11}) P_1\, p_1(\underline{x}) + (C_{21}-C_{22}) P_2\, p_2(\underline{x})\}\,d\underline{x}$$

$$P_1 + P_2 = 1 \longrightarrow P_2 = 1 - P_1$$

$$r = C_{22} + \underbrace{(C_{21}-C_{22})\int_{R_1} p_2(\underline{x})\,d\underline{x}}_{cte} + P_1\Big[\underbrace{(C_{11}-C_{22}) + (C_{12}-C_{11})\int_{R_2} p_1(\underline{x})\,d\underline{x} - (C_{21}-C_{22})\int_{R_1} p_2(\underline{x})\,d\underline{x}}_{cte}\Big]$$

$r = K_1 + P_1 K_2$



$P_1 = 0$

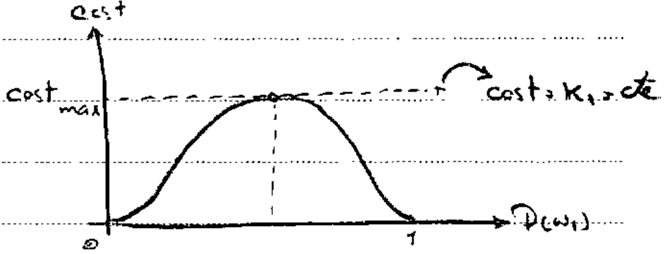
$P_1 = 1$ $\Rightarrow$ cost = 0

چرا تابع cost - P(ω₁) یک تابع مقعر به پایین است؟ با اثبات ریاضی

• برای کم کردن اثر P روی cost، باید بیشینه را کم کرد ← minmax

$r = K_1 + P_1 . K_2 \xrightarrow{\text{minimize}} K_2 = 0$

یعنی خط r مستقل از P_1 باشد



$$(C_{11} - C_{22}) + (C_{12} - C_{11}) \int_{R_2} P_1(x)\,dx = (C_{21} - C_{22}) \int_{R_1} P_2(x)\,dx$$

• حالات خاص

* هزینه تصمیم گیری صفر باشد و خطاها یکسان باشد

$C_{11} = C_{22} = 0$

$C_{21} = C_{12}$

$$\Rightarrow \int_{R_2} P_1(x)\,dx = \int_{R_1} P_2(x)\,dx \Rightarrow \mathcal{E}_1 = \mathcal{E}_2$$

Subject:
Year. Month. Date. 24

Reciever Operating Characteristic (ROC) *

$$\varepsilon_1 = \int_{R_2} p_1(\underline{x}) \cdot d\underline{x} \quad :$$

$$\varepsilon_2 = \int_{R_1} p_2(\underline{x}) \cdot d\underline{x}$$

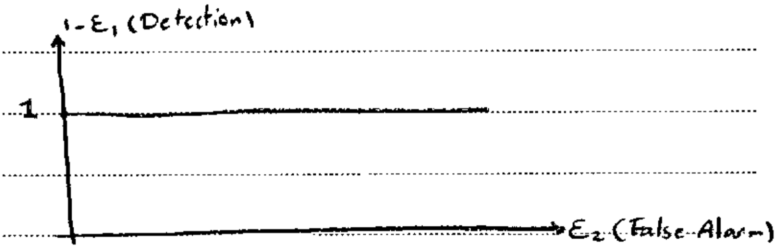
- $\varepsilon_i \longrightarrow$ Error Probability Type (i)

- ω_1 : Positive Class ⟶ False Negative Rate (FNR)
ω_2 : Negative Class ⟶ False Positive Rate (FPR)

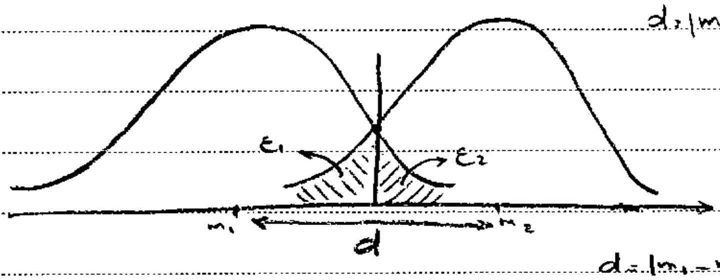
- Radar

ω_1 : Target ⟶ Missed Detection
ω_2 : Noise ⟶ False Alarm

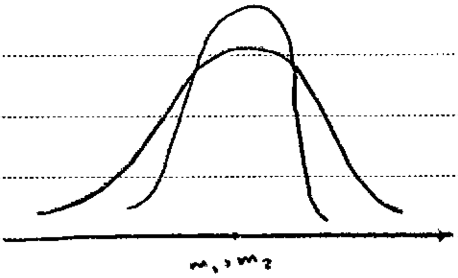
Detection : $1 - \varepsilon_1$
مطلوب کشف 100% است

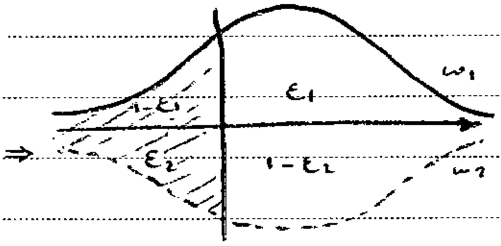


- میانگین متفاوت $d = |m_1 - m_2|$



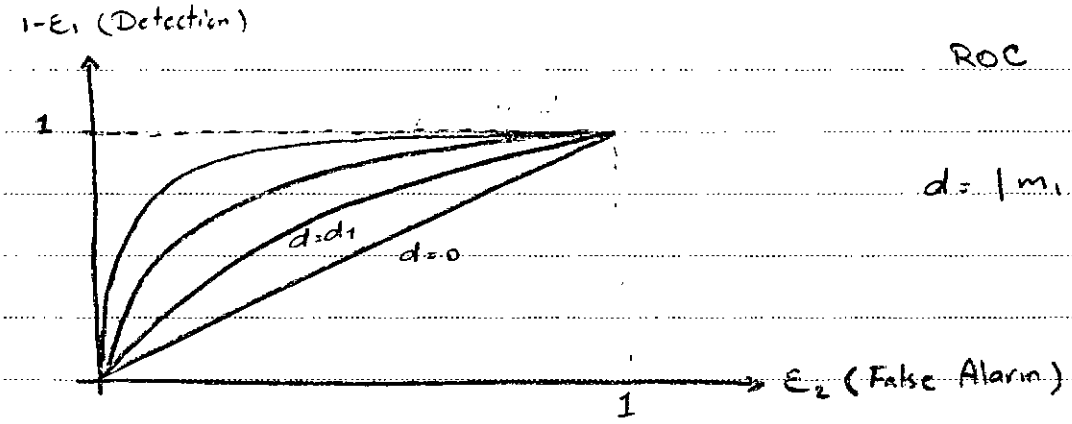
- میانگین یکسان $d = |m_1 - m_2| = 0$

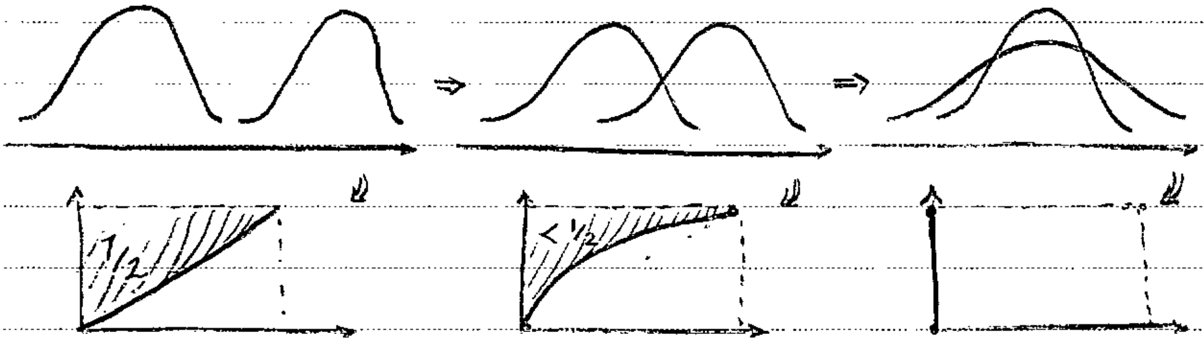




- منحنی ROC

$d = |m_1 - m_2|$





Multi - hypothesis Testing *

M-class problem

$q_1(\underline{x}), \ldots, q_M(\underline{x})$

$\underline{x} \in \omega_i \qquad q_i(\underline{x}) > q_j(\underline{x})$

$$q_i(\underline{x}) = \max_{\forall j} q_j(\underline{x}) \qquad \forall j \neq i$$

$C_{ij} \qquad$ i: true class $\qquad$ j: decision

$$r = \sum_{j=1}^{M} \sum_{i=1}^{M} \int_{R_j} C_{ij} P_i p_i(\underline{x}) . dx$$

Single - hypothesis Testing *

. مشخصات آماری تنها یک کلاس موجود است.

☑ در نظارت تصویری، تشخیص تانک دشمن است.

Target Class: t

$\underline{m}_t \quad , \quad [\Sigma]_t$

$d(\underline{x})^2 = (\underline{x} - \underline{m}_t)^T [\Sigma]_t^{-1} (\underline{x} - \underline{m}_t)$ ← فاصله ماهالانوبیس

اگر $\underline{X}$ متغیر تصادفی باشد، فاصله نیز متغیر تصادفی می‌شود.

$D^2 = (\underline{X} - \underline{m}_t)^T [\Sigma]_t^{-1} (\underline{X} - \underline{m}_t)$

با یک تبدیل Whitening

$\underline{Z} = [a]^T (\underline{X} - \underline{m}_t)$

$$\Rightarrow \quad E\{\underline{Z}\} = \underline{0} \quad , \quad \mathrm{cov}\{\underline{Z}\} = [I]$$

$$D^2 = \underline{Z}^T \underline{Z}$$

$$\underline{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_n \end{bmatrix} \Rightarrow D^2 = \sum_{i=1}^{n} Z_i^2 \longrightarrow Z_i \text{ are uncorrelated}$$

$$E\{D^2\} = E\{\underline{Z}^T \underline{Z}\} = E\{\sum_{i=1}^{n} Z_i^2\} = \sum E\{\overset{1}{Z_i^2}\} = n$$

$$\mathrm{var}\{D^2\} = E\{(D^2)^2\} - E^2\{D^2\}$$

$$= \sum_i \{Z_i^4\} + \sum_{\substack{i \\ i \neq j}} \sum_j E\{Z_i^2 Z_j^2\} - E^2\{D^2\}$$

if Z_i & Z_j are independant

$$\mathrm{var}\{D^2\} = n E\{Z_i^4\} + n(n-1) \underbrace{E\{Z_i^2\}}_{1} \underbrace{E\{Z_i^2\}}_{1} - \underbrace{E^2\{D^2\}}_{n^2}$$

$$= n E\{Z_i^2\} - n E^2\underbrace{\{Z_i^2\}}_{1} = n\gamma$$

$$\gamma = E\{Z_i^4\} - 1$$

• برای توزیع نرمال Z_i

$$E\{Z_i^4\} = 3 \underbrace{E^2\{Z_i^2\}}_{1} = 3 \Longrightarrow \gamma = 2$$

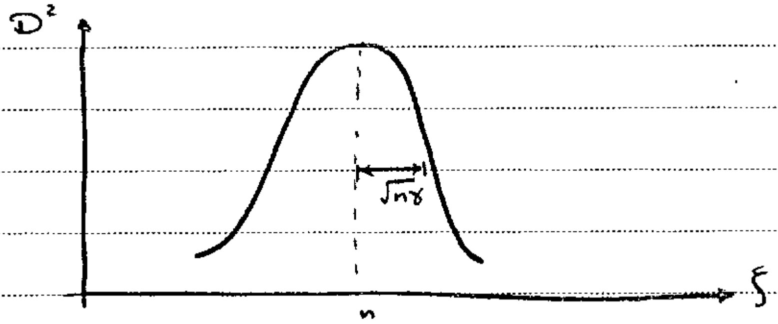
پس برای تست نرمال بودن توزیع داده ها ⟵ $\gamma \overset{?}{=} 2$

$$E\{D^2 \mid \omega_t\} = n$$

$$\mathrm{var}\{D^2 \mid \omega_t\} = n\gamma \quad \text{where} \quad \gamma = E\{Z_i^4\} - 1$$

$$D^2 = \underline{Z}^T Z = \sum_{i=1}^{n} Z_i^2$$

$n \gg 0$ $\xrightarrow{\text{Central Limit Theorem}}$ توزیع D^2 به سمت توزیع نرمال می‌رود



$E\{D^2\} = n$
$\sigma^2\{D^2\} = n\gamma$

Performance of a Single Hypothesis Test

$\underline{m}_1 \equiv \underline{m}_t = 0 \qquad [\Sigma]_1 = [\Sigma]_t = [I]$

$\underline{m}_2 \equiv \underline{m} =$ non-target class $\qquad [\Sigma]_2 = [\lambda] = \mathrm{diag}\{\lambda_1, \dots, \lambda_n\}$

هر مسئله را با قطری کردن همزمان می‌توان به این شرایط رساند

$$D^2 = (\underline{X} - \underline{m}_t)^T [\Sigma]_t^{-1} (\underline{X} - \underline{m}_t) \curvearrowright$$ فاصله از کلاس هدف

$$= Z^T Z$$
$$= \sum_{i=1}^{n} Z_i^2$$

$E\{D^2 | \omega_t\} = n$
$\mathrm{var}\{D^2 | \omega_t\} = n\gamma$

$$D^2 = \underline{X}^T \underline{X} = (\underline{X} - \underline{m} + \underline{m})^T (\underline{X} - \underline{m} + \underline{m})$$
$$= (\underline{X} - \underline{m})^T (\underline{X} - \underline{m}) + 2\underline{m}^T (\underline{X} - \underline{m}) + \underline{m}^T \underline{m}$$
$$= tr\left[(\underline{X} - \underline{m})^T (\underline{X} - \underline{m})\right] + 2\underline{m}^T (\underline{X} - \underline{m}) + \underline{m}^T \underline{m}$$

$$E\{D^2|\omega_2\} = tr\left[\underbrace{E\{(\underline{X}-\underline{m})(\underline{X}-\underline{m})^T\}}_{}\,|\omega_2\right] + 2m^T E\{\underbrace{(\underline{X}-\underline{m})|\omega_2}_{}\} + \underline{m}^T\underline{m}$$

$$= \boxed{\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n} m_i^2}$$

$$var\{D^2|\omega_2\} = \underbrace{E\{(D^2)^2|\omega_2\}}_{\swarrow} - E^2\{D^2|\omega_2\}$$

$$E\{(D^2)^2|\omega_2\} = E\{(\underline{X}-\underline{m})^T(\underline{X}-\underline{m})(\underline{X}-\underline{m})^T(\underline{X}-\underline{m})\,|\omega_2\}$$
$$+ 4\,\underline{m}^T E\{(\underline{X}-\underline{m})^T(\underline{X}-\underline{m})\,|\omega_2\}\,\underline{m}$$
$$+ 2\,E\{(\underline{X}-\underline{m})^T(\underline{X}-\underline{m})\,|\omega_2\}\,\underline{m}^T\underline{m}$$
$$+ (\underline{m}^T\underline{m})^2$$
$$= 3\sum_i \lambda_i^2 + \sum_i\sum_j \lambda_i\lambda_j \quad (i\neq j)$$
$$+ 4\sum_i \lambda_i m_i^2$$
$$+ (\sum m_i^2)^2 + 2(\sum_i \lambda_i)(\sum m_i^2)$$

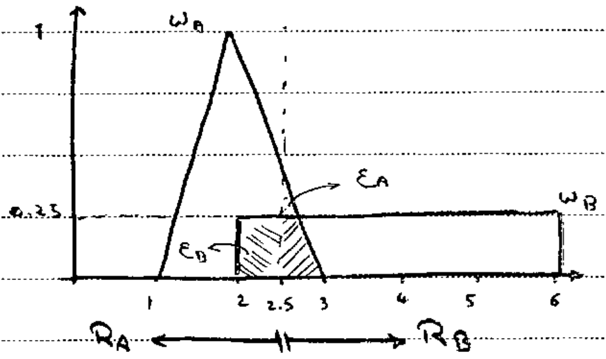
$$var\{D^2|\omega_2\} = \boxed{2\sum_{i=1}^{n}\lambda_i^2 + 4\sum_{i=1}^{n}\lambda_i m_i^2}$$

ε بدست آمده حاصل این نگاشت از n بعد (ε_x) به یک بعد (ε_d) بسیار بالاتر است.

Estimation Of Error Rates *

$$\varepsilon = P_1 \varepsilon_1 + P_2 \varepsilon_2$$
$$= P_1 \int_{R_2} p_1(\underline{x}) \cdot d\underline{x} + P_2 \int_{R_1} p_2(\underline{x}) \cdot d\underline{x}$$

☑



$h(\underline{x}) = x - 2.5 = 0$

$\varepsilon_B = \frac{1}{8}$

$\varepsilon_A = 0.125$

Sample Counting *

N Sample
$$\begin{cases} N_1 \in \omega_1 \\ N_2 \in \omega_2 \\ N_1 + N_2 = N \end{cases}$$

$h(\underline{x}) \rightarrow$ معلوم

- تعداد داده زیاد

Training Sample → آموزش سیستم و تابع تمایز

Test Sample → ارزیابی سیستم

تعداد داده‌های خطا را شمارش می‌کنیم.

اگر آزمایش به دفعات بالا رسد، میانگین تخمین به مقدار واقعی متمایل می‌شود

$$N \begin{cases} N-1 : \text{training} \\ 1 : \text{test} \end{cases}$$ ☑

این آزمایش را N-1 بار تکرار می‌کنیم ← روش Leave-one-out

Linear Boundaries *

ω_1 & ω_2 are Normally Distributed

$[\Sigma]_1 = [\Sigma]_2 = [\Sigma]$

$$H \equiv h(\underline{x}) = (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} \underline{X} + \frac{1}{2} (\underline{m}_1^T [\Sigma]^{-1} \underline{m}_1 - \underline{m}_2^T [\Sigma] \underline{m}_2) \overset{\omega_1}{\underset{\omega_2}{\lessgtr}} \ln \frac{P_1}{P_2}$$

خطی ↗

$*\ E\{H(\underline{x}) \mid \omega_i\} = \eta_i$

$$\eta_i = E\{H(\underline{x}) \mid \omega_i\}$$
$$= (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} \underline{m}_i + \frac{1}{2} \underbrace{(\underline{m}_1^T [\Sigma]^{-1} \underline{m}_1 - \underline{m}_2^T [\Sigma]^{-1} \underline{m}_2)}_{(\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} (\underline{m}_2 - \underline{m}_1)} \overset{\omega_1}{\underset{\omega_2}{\gtrless}} \ln \frac{P_1}{P_2}$$

$$\eta_1 = -\frac{1}{2} (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} (\underline{m}_2 - \underline{m}_1) = \underline{-\eta}\,\Big|$$

$$\eta_2 = +\frac{1}{2} (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} (\underline{m}_2 - \underline{m}_1) = \underline{+\eta}\,\Big|$$

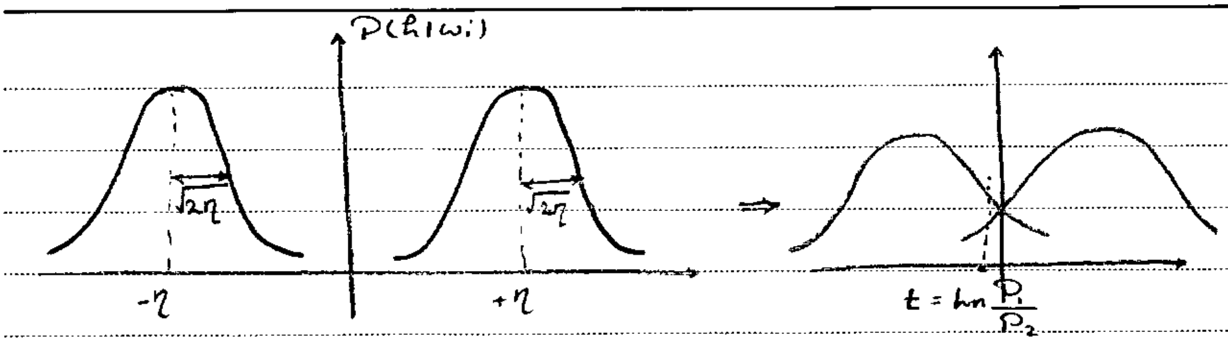
$*\ Var\{H(x) \mid \omega_i\} = \sigma_i^2 = E\{(H - \eta_i)^2 \mid \omega_i\}$

$$\sigma_i^2 = E\left\{\left((\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} X - \frac{1}{2} (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} (\underline{m}_2 - \underline{m}_1) - \eta_i\right)^2\right\}$$

$$= E\left\{\left[(\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} (\underline{X} - \underline{m}_i)\right]^2 \mid \omega_i\right\}$$

$$= (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} \underbrace{E\{(\underline{X} - \underline{m}_i)(\underline{X} - \underline{m}_i)^T \mid \omega_i\}}_{[\Sigma]_i \rightarrow [\Sigma]} [\Sigma]^{-1} (\underline{m}_2 - \underline{m}_1)$$

$$= (\underline{m}_2 - \underline{m}_1)^T [\Sigma]^{-1} (\underline{m}_2 - \underline{m}_1) = \underline{2\eta}\,\Big|$$



$$\varepsilon_1 = \int_t^{+\infty} P(h|\omega_1)\, dh \Rightarrow \varepsilon_1 = \int_{\frac{t+\eta}{\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\xi/2}\, d\xi$$

$$\varepsilon_2 = \int_{-\infty}^{t} P(h|\omega_2)\, dh \Rightarrow \varepsilon_2 = \int_{-\infty}^{\frac{t+\eta}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\xi/2}\, d\xi$$

*

$h(\underline{x}) \longrightarrow$ Quadratic function

$[\Sigma]_1 \,,\, [\Sigma]_2$

$$H(\underline{x}) = \underbrace{\frac{1}{2}(\underline{X}-\underline{m}_1)^T[\Sigma]_1^{-1}(\underline{X}-\underline{m}_1)}_{} - \underbrace{\frac{1}{2}(\underline{X}-\underline{m}_1)^T[\Sigma]_2^{-1}(\underline{X}-\underline{m}_2)}_{B} + \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} - t \overset{\omega_2}{\underset{\omega_1}{\gtrless}} 0$$

$$\underbrace{\mathrm{trace}\,([\Sigma]_1^{-1}(\underline{X}-\underline{m}_1)'(\underline{X}-\underline{m}_1)^T}_{A}$$

$$E\{A\} = \mathrm{trace}\ [\Sigma]_1^{-1}\ E\{(\underline{X}-\underline{m}_1)(\underline{X}-\underline{m}_1)^T$$

$$E\{B\} = -\frac{1}{2}\, E\,\{(\underline{X}-\underline{m}_2)^T\,[\Sigma]_2^{-1}\,(\underline{X}-\underline{m}_2)$$

$$= -\frac{1}{2}\, E\,\{((x-m_1)-(m_2-m_1))^T\,[\Sigma]_2^{-1}\,((\underline{X}-\underline{m}_1)-(\underline{m}_2-\underline{m}_1))$$

$$= -\frac{1}{2}\, E\left\{\mathrm{trace}\,\{[\Sigma]_2^{-1}\,[(\underline{X}-\underline{m}_1)(\underline{X}-\underline{m}_1)^T - (\underline{m}_2-\underline{m}_1)(\overset{0}{\cancel{\underline{X}}}-\underline{m}_1)^T - (X-\cancel{m_1})(\overset{0}{m_2}-m_1)^T + (\underline{m}_2-\underline{m}_1)(\underline{m}_2-\underline{m}_1)^T]\}\right\}$$

$$= -\frac{1}{2}\,\text{trace}\left[[\Sigma]_2^{-1}[\Sigma]_1 + (m_2-m_1)(m_2-m_1)^T\right]$$

$$\eta_1 = E\{H(\underline{x})\,|\,\omega_1\} = A + B = \frac{1}{2}\left\{\text{trace}\left\{[I] - [\Sigma]_2^{-1}[\Sigma]_1\right\}\right.$$

$$-\frac{1}{2}(m_2-m_1)^T[\Sigma]_2^{-1}(m_2-m_1)$$

$$+\frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} - t$$

$$\eta_2 = E\{H(\underline{x})\,|\,\omega_2\} = -\eta_1$$

با قطری کردن همزمان در ماتریس Σ_1 و Σ_2 داریم:

$$E\{H(\underline{x})\,|\,\omega_1\} = \frac{1}{2}\sum_{i=1}^{n}\left[\left(1-\frac{1}{\lambda_i}\right) - \frac{(d_{2i}-d_{1i})^2}{\lambda_i} + \ln\frac{1}{\lambda_i}\right] - t$$

$$\text{var}\{H(\underline{x})\,|\,\omega_1\} = E\{(H(\underline{x})-\eta_1)^2\,|\,\omega_1\} = \frac{1}{2}\sum_{i=1}^{n}\left[\left(1-\frac{1}{\lambda_i}\right) + 2\frac{(d_{2i}-d_{1i})^2}{\lambda_i^2}\right]$$

در کل برابر است با:

$$\text{var}\{H(\underline{x})\,|\,\omega_1\} = \frac{1}{2}\,\text{tr}\left\{([I]-[\Sigma]_2^{-1}[\Sigma]_1)^2 + (\underline{m}_2-\underline{m}_1)^T[\Sigma]_2^{-1}[\Sigma]_1[\Sigma]_2^{-1}(\underline{m}_2-\underline{m}_1)\right\}$$

Upper Bounds on the Bayes Error *

Charnoff Bound.

Bayes Error : $\varepsilon = \int \min\left[P_1 p_1(x),\ P_2 p_2(x)\right].d\underline{x}$

Subject:
Year. Month. Date. 29

$a, b \geqslant 0$

$\min[a,b] \leqslant a^s b^{1-s}$, $0<s<1$

اثبات:

$a^s b^{1-s} = a \times \left(\frac{b}{a}\right)^{1-s}$

$1-s > 0$

فرض $\min[a,b] = a \Rightarrow \left(\frac{b}{a}\right) \geqslant 1$

$\Rightarrow a \times \left(\frac{b}{a}\right)^{1-s} \geqslant a = \min[a,b]$

Chernoff Bound

$0<s<1$

$$\mathcal{E}_u = P_1^{\,s} \cdot P_2^{(1-s)} \int p_1^{s}(\underline{x}) \cdot p_2^{(1-s)}(\underline{x})\, d\underline{x}$$

• حالت خاص:

$\omega_1 : N(\underline{m}_1, [\Sigma]_1)$

$\omega_2 : N(\underline{m}_2, [\Sigma]_2)$

$$\int p_1^{s}(\underline{x})\, p_2^{(1-s)}(\underline{x})\, d\underline{x} = e^{-\mu(s)}$$

فاصله Chernoff :

$$\mu(s) = \frac{s(1-s)}{2} (\underline{m}_2 - \underline{m}_1)^T \Big[s[\Sigma]_1 + (1-s)[\Sigma_2] \Big]^{-1} (\underline{m}_2 - \underline{m}_1)$$

فاصله میانگین ها را [illegible]

$$+ \frac{1}{2} \ln \frac{|s[\Sigma]_1 + (1-s)[\Sigma]_2|}{|[\Sigma]_1|^{s} \, |[\Sigma]_2|^{(1-s)}}$$

فاصله برای ماتریس کواریانس

• حالت خاص فاصله chernoff :

$s = \frac{1}{2} \longrightarrow$ Bhattacharyya Distance

$$\mu\left(\frac{1}{2}\right) = \frac{1}{8} (m_2 - m_1) \left[\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2 \right]^{-1} (m_2 - m_1) + \frac{1}{2} \ln \frac{\left|\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2\right|}{\sqrt{|\Sigma_1|\,|\Sigma_2|}}$$

شاخص فاصله Bhattacharyya معیار خوبی برای مقایسه دو توزیع است و میزان نزدیک و دور بودن آنها را تعیین می‌کند.

$S \Rightarrow \mu(s) \uparrow \Rightarrow \varepsilon_u \downarrow$

فصل سوم ← تمرینات کامپیوتری 1, 2, 3, 4, 6, 8

مسائل 1, 2, 3, 4 (a ~ c), 5, 6, 7, 10

Hand Out

تاریخ تحویل : ۸۷ ,۸,۲۱

* روشهای محاسبه خطا

- $P_i(\underline{x}) \longrightarrow \varepsilon = P_1 \varepsilon_1 + P_2 \varepsilon_2$
- $P_i(h) \longrightarrow$ mapping $nD \rightarrow 1D$
- Approximation ε
- Upper bound ⟶ chernoff

$$\varepsilon = \int_{-\infty}^{+\infty} \min \left[P_1 \, p_1(\underline{x}) , P_2 \, p_2(\underline{x})\right] d\underline{x} \leq P_1^s P_2^{(1-s)} \int_{-\infty}^{+\infty} p_1^s(\underline{x}) \, p_2^{(1-s)}(\underline{x}) \, d\underline{x}$$

$\omega_1 : N(m_1, [\Sigma]_1)$

$\omega_2 : N(m_2, [\Sigma]_2)$

$$\int p_1^s(\underline{x}) \, p_2^{(1-s)}(\underline{x}) \, d\underline{x} = e^{-\mu(s)} \longrightarrow \mu(s) = \text{chernoff distance}$$

اثبات :

$$\left.\begin{array}{l} [\Sigma]_1 \longrightarrow [I] \\ [\Sigma]_2 \longrightarrow [\lambda] \\ \underline{Y} = [a]^T \underline{x} \end{array}\right| \Rightarrow \text{تعریف میانگین} \Rightarrow \begin{cases} m_1 = \underline{d}_1 \\ m_2 = \underline{d}_2 \end{cases}$$

$$\underline{d}_1 = [\, d_{11} \;\; d_{12} \; \dots \; \underbrace{d_{1\ell}}_{\ell^{th} \text{ element of } \underline{d}_1} \; \dots \; d_{1n} \,]^T$$

$$A = \int P_1^{(1-s)}(\underline{x}) \cdot P_2^{(s)}(\underline{x}) \cdot d\underline{x} = \int \frac{\exp\left[-\frac{1}{2}\{(1-s)(\underline{y}-\underline{d}_1)^T(\underline{y}-\underline{d}_1) + s(\underline{y}-\underline{d}_2)^T[\lambda]^{-1}(\underline{y}-\underline{d}_2)\}\right]}{(2\pi)^{n/2}\,|[I]|^{(1-s)/2} \cdot |[\lambda]|^{s/2}} d\underline{y}$$

$$A = \prod_{\ell=1}^{n} \int_{-\infty}^{+\infty} \frac{1}{\lambda_\ell^{s/2}\,(2\pi)^{1/2}} \cdot \exp\left[\underbrace{-\frac{1}{2}\left\{(1-s)(y_\ell - d_{1\ell})^2 + \frac{s}{\lambda_\ell}(y_\ell - d_{2\ell})^2\right\}}_{\text{کامل کردن مربع}}\right] \cdot d\underline{y}$$

$$A = \prod \frac{1}{\lambda_\ell^{s/2}\left\{(1-s)+\frac{s}{\lambda_\ell}\right\}^{1/2}} \times \exp\left\{-\frac{s(1-s)}{2} \cdot \frac{(d_{2\ell}-d_{1\ell})^2}{s+(1-s)\lambda_\ell}\right\}$$

$$\times \int_{-\infty}^{+\infty} \frac{\left\{(1-s)+\frac{s}{\lambda_\ell}\right\}^2}{(2\pi)^{1/2}} \exp\left[-\underbrace{\frac{\{(1-s)+s/\lambda_\ell\}}{2}}_{\alpha} \left\{y_\ell - d_\ell - \frac{(\frac{s}{\lambda_\ell}(d_{2\ell}-d_{\ell})}{(1-s)+s/\lambda_\ell}\right\}^2\right] d\underline{y}$$

$$\int_{-\infty}^{+\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \longleftarrow \alpha$$

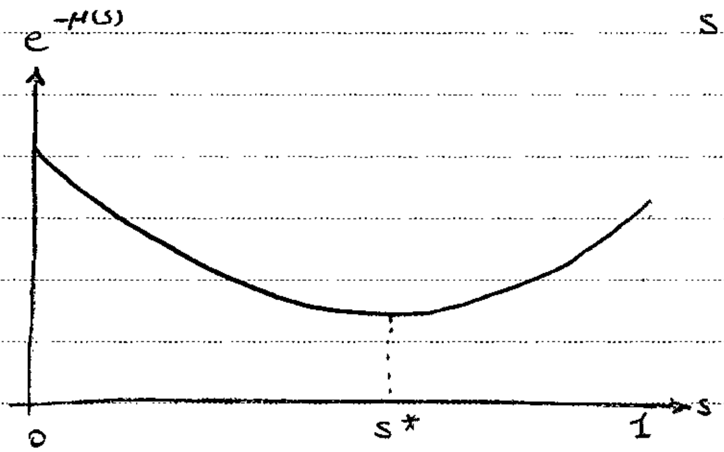
$$A = \prod \frac{\lambda_\ell^{(1-s)/2}}{\{s+(1-s)\lambda_\ell\}^{1/2}} \exp\left[-\frac{s(1-s)}{2} \cdot \frac{(d_{2\ell}-d_{1\ell})^2}{s+(1-s)\lambda_\ell}\right]$$

حال معادله بسط داده شده را به حالت ماتریسی بر می گردانیم.

$$A = \frac{|[I]|^{s/2} \quad |[\lambda]|^{(1-s)/2}}{|s[I]+(1-s)[\lambda]|^{1/2}} \cdot \exp\left[-\frac{s(1-s)}{2}(\underline{d}_2-\underline{d}_1)^T\{s[I]+(1-s)[\lambda]\}^{-1}(\underline{d}_2-\underline{d}_1)\right]$$

حال از تغییر Y به X بر می گردیم و فرمول ثابت می شود.

• پیدا کردن مقدار بهینه s



* روش عددی
* روش آنالیزی

Bhattacharyya Bound ⟶ $s = 1/2$ •

$$\varepsilon_u = \sqrt{P_1P_2}\int_{-\infty}^{+\infty}\sqrt{p_1(\underline{x})\,p_2(\underline{x})}\;d\,\underline{x}) = \sqrt{P_1P_2}\;e^{-\mu(\frac{1}{2})}$$

$$\mu(\tfrac{1}{2}) = \frac{1}{8}(m_2-m_1)^T\left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1}(m_2-m_1) + \frac{1}{2}\ln\frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1|\times|\Sigma_2|}}$$

• فاصله Bhattacharyya : میزان شبیه بودن دو توزیع
* قسمت اول : پراکندگی حول میانگین
* قسمت دوم : واریانس حول میانگین

* برای انتخاب ویژگی ها از این فاصله استفاده می شود تا ویژگیها از هم دور باشند
پس feature ای انتخاب می شود تا این فاصله بیشینه شود.

☑

$\omega_1 : N_x(\underline{0}, [I])$

$\omega_2 : N_x(\underline{0}, [\lambda])$

B.D.:

$\mu(\frac{1}{2}) = \frac{1}{2} \sum_{i=1}^{n} Ln \frac{1+\lambda i}{2\sqrt{\lambda i}}$

$[\Sigma_2] = diag\{\lambda_1, \ldots, \lambda_n\} \longrightarrow \lambda_i > 0 \longrightarrow \frac{1+\lambda i}{2\sqrt{\lambda i}} \geqslant 1$

$\lambda i = 1 : \omega_1 = \omega_2$!!!

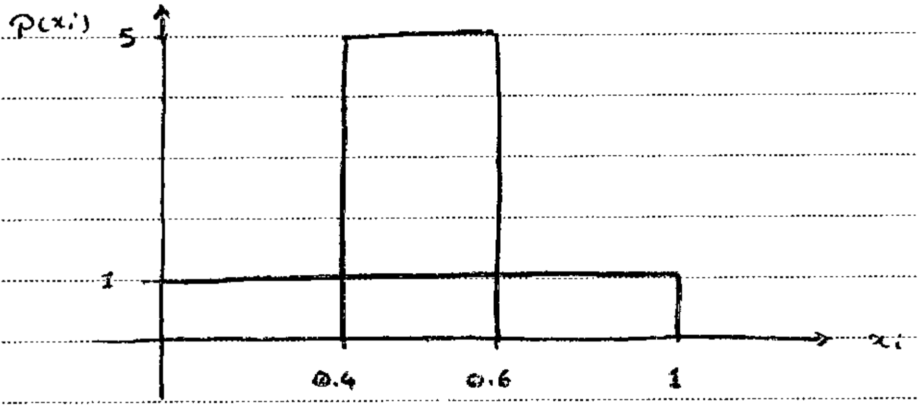
$\lambda i \neq 0 : n \longrightarrow +\infty \Longrightarrow \mu(\frac{1}{2}) \longrightarrow \infty \Longrightarrow \varepsilon_u = e^{-\mu(\frac{1}{2})} \longrightarrow 0$

n-D ☑

$X_i \longrightarrow$ iid = independent & identically distributed

$\omega_1 = P_1(\underline{x}) \longrightarrow$ uniformally distributed $[0.4, 0.6]$

$\omega_2 = P_2(\underline{x}) \longrightarrow$ uniformally distributed $[0, 1]$



B.B.: $\varepsilon_u = \sqrt{P_1P_2} \int_{-\infty}^{+\infty} \sqrt{P_1(\underline{x}).P_2(\underline{x})} .d\underline{x} \xRightarrow{iid} \varepsilon_u = \sqrt{P_1P_2} \prod_{i=1}^{n} \int \sqrt{P_1(x_i).P_2(x_i)} dx_i$

$\varepsilon_u = \sqrt{P_1P_2} \prod_{i=1}^{n} \int_{0.4}^{0.6} \sqrt{(5)(1)} .dx_i = \sqrt{P_1P_2} \prod (\sqrt{5}(0.2)) = \sqrt{P_1P_2}(0.447)^n$

$n \longrightarrow +\infty \Longrightarrow \varepsilon_u \longrightarrow 0$

• با اضافه کردن Feature ؛

← اضافه کردن dimentionality

← $E_u \Downarrow$

← بهبود کارایی (خطای کم)

ولیکن پس از یک حد، کارایی دیگر بالا نمی‌رود که شناسایی الگو تنها در عمل صورت می‌نماید و به آن Curse of Dimentionality می‌گویند.

• انواع دیگری از حد بالا وجود دارد:

$\min [a,b] \leqslant a^{s} b^{(1-s)}$ chernoff *

$\min [a,b] \leqslant \sqrt{ab}$ bhattacharyya *

$\min [a,b] \leqslant \frac{2ab}{(a+b)}$ assymptotic nearest neighbor *

$\min [a,b] \leqslant \underline{\frac{2ab}{(a+b)}} \leqslant \sqrt{ab}$ ← ترتیب تنگ‌تر بودن از BD

* Charasteristic Function of Random Vector

$$F(\omega) = E\left\{\exp(-j\omega X)\right\}$$

• 1-D

$$F(\omega) = \int_{-\infty}^{+\infty} P_x(x) . \exp(+j\omega X) . d_x$$

عملکردی شبیه تبدیل فوریه دارد.

Parametric Classifiers *

• با دانستن تابع توزیع داده‌ها از روش‌های فصل قبل استفاده می‌شود.
در این حالت میانگین و کواریانس موجود است.

• در این فصل n داده موجود است.

$\underline{x}_i^K$ $i=1,\dots,N$
$K=1,\dots,M$

$h(\underline{x})$ → desired parametric form of classifier

☑ $h(\underline{x})$ ~ Linear Classifier

$$h(\underline{x}) = \underline{v}^T \underline{x}' + v_0 = \underline{w}^T \underline{x}$$

v_0 ↳ threshold → to be learned

$\underline{v}$ → weight vector → n-D

$$h(\underline{x}) = \underline{v}^T \underline{x} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} -v_0$$

$$\underline{x} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \underline{w} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_n \end{bmatrix}$$ — augmented weight vector

☑ $h(\underline{x})$ ~ Non-Linear

Goal: define $\underline{v}^T$ and v_0

INTRO

☑ Bayes Classifier

2- Class

$\underline{m}_i \, , [\Sigma]_i \qquad i = 1, 2$

$$* \; h(\underline{x}) = \frac{1}{2}(\underline{x}-\underline{m}_1)^T [\Sigma]_1^{-1} (\underline{x}-\underline{m}_1) - \frac{1}{2}(\underline{x}-\underline{m}_2) [\Sigma]_2^{-1} (\underline{x}-\underline{m}_2) + \frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} \ln \frac{P_1}{P_2}$$

$* \text{ if } \Sigma_1 = \Sigma_2 = \Sigma$

$$h(\underline{x}) = (\underline{m}_2-\underline{m}_1)^T \Sigma^{-1} \underline{x} + \frac{1}{2}(\underline{m}_1^T \Sigma^{-1} \underline{m}_1 - \underline{m}_2^T \Sigma^{-1} \underline{m}_2) \gtrless \ln \frac{P_1}{P_2}$$

$$\underline{v}^T = (\underline{m}_2 - \underline{m}_1)^T \Sigma^{-1}$$

$$v_0 = \frac{1}{2}(\underline{m}_1^T \Sigma \underline{m}_1 - \underline{m}_2^T \Sigma^{-1} \underline{m}_2) - \ln \frac{P_1}{P_2}$$

$* \text{ if } \Sigma_1 = \Sigma_2 = [I]$

$$h(\underline{x}) = (\underline{m}_2-\underline{m}_1)^T \underline{x} + \frac{1}{2}(\underline{m}_1^T \underline{m}_1 - \underline{m}_2^T \underline{m}_2) \gtrless \ln \frac{P_1}{P_2}$$

$$\underline{v}^T = (\underline{m}_2 - \underline{m}_1)$$

$$v_0 = \frac{1}{2}(\underline{m}_1^T \underline{m}_1 - \underline{m}_2^T \underline{m}_2) - \ln \frac{P_1}{P_2}$$

Correlation

$$x_3[n] = x_1[n] \circledast x_2[n] = \sum_m x_1[m]\, x_2[m+n]$$

$$x_3[0] = \sum_m x_1[m]\, x_2[m] = x_1^T x_2 \rightarrow \text{correlation at origin}$$

برای همین از $\Sigma_1 = \Sigma_2 = [I]$ با Bayes Clas. به Correlation Clas تبدیل می‌شود

در Optical Pattern Recog. معمولاً با Correlation Classifier انجام می‌شود

سرعت پردازش به اندازه سرعت نور است.

*

Bayes Classifier: $\Sigma_1 = \Sigma_2 = [I]$

$$h(\underline{x}) = (\underline{m}_2 - \underline{m}_1)^T \underline{x} + \frac{1}{2} (\underline{m}_1^T \underline{m}_1 - \underline{m}_2^T \underline{m}_2) \gtrless \ln \frac{P_1}{P_2}$$

حال اگر به آن $\underline{x}^T \underline{x}$ اضافه کم کنیم:

$$h(\underline{x}) = \underbrace{(\underline{x}^T \underline{x} - 2 \underline{m}_1^T \underline{x} + \underline{m}_1^T \underline{m}_1)}_{\|x - m_1\|^2} - \underbrace{(\underline{x}^T \underline{x} - 2 \underline{m}_2^T \underline{x} + \underline{m}_2^T \underline{m}_2)}_{\|\underline{x} - m_2\|^2} \overset{\omega_2}{\underset{\omega_1}{\gtrless}} 2 \ln \frac{P_1}{P_2}$$

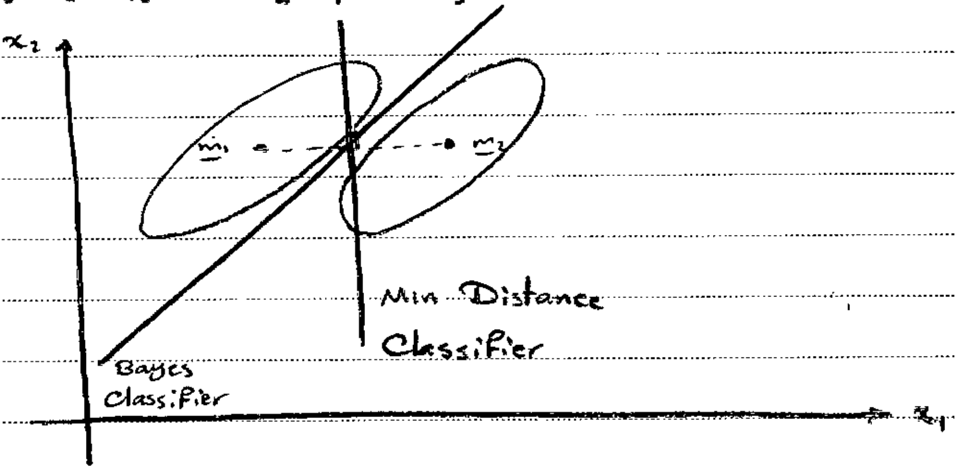
$$\text{Bayes Classifier} \xrightarrow{\Sigma_1 = \Sigma_2 = [I]} \text{Min Distance Classifier}$$

نتیجه

$$\|\underline{x} - \underline{m}_1\|^2 - \|\underline{x} - \underline{m}_2\|^2 \overset{\omega_2}{\underset{\omega_1}{\gtrless}} 2 \ln \frac{P_1}{P_2}$$

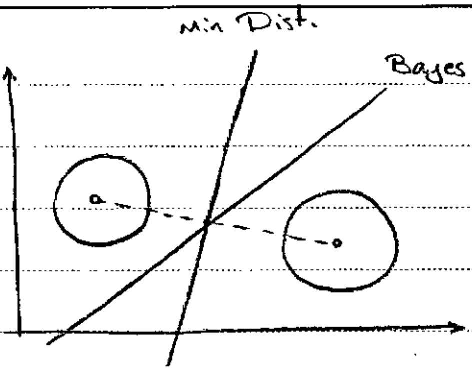
خط عمود منصف

* $[\Sigma] = [\Sigma_1] = [\Sigma]_2 \neq [I]$



Whitening Transform

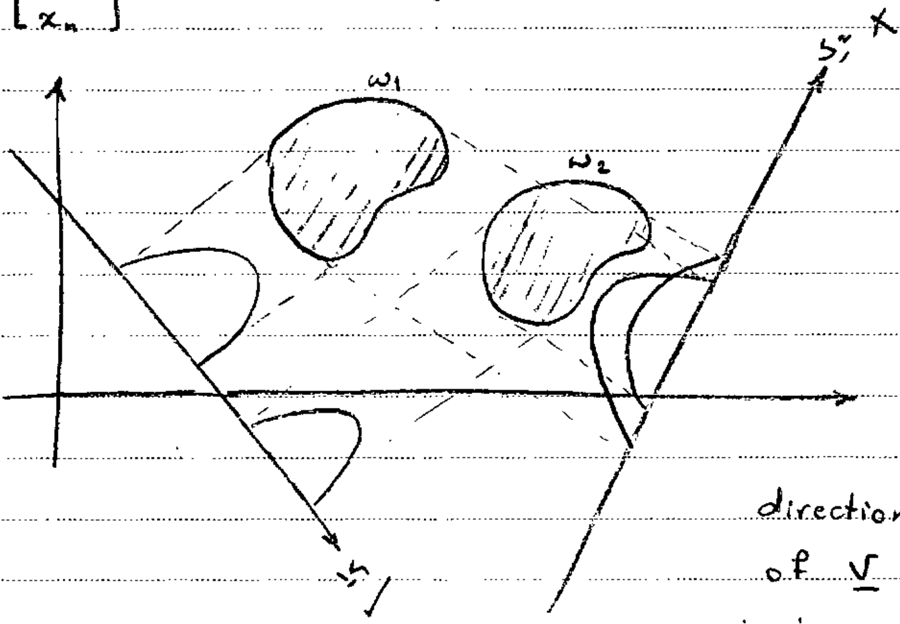
$$[\Sigma] \longrightarrow [I]$$



$P(\underline{x}) \sim$ مجهول

$h(\underline{x}) = \underline{v}^T \underline{x} + v_0 \sim$ مطلوب

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \Longrightarrow \begin{array}{l} y = \underline{v}^T x_0 \\ y \gtrless -v_0 \end{array}$$



direction
of $\underline{v}$
is important

$h(\underline{x}) = \underline{v}^T \underline{x} + v_0$

$H(\underline{x}) = v_1 X_1 + \cdots + v_n X_n + v_0$

طبق قضیه حد مرکزی این بردار تصادفی توزیع نرمال دارد که داریم:

$$\eta_i = E\{H|\omega_i\}$$
$$\sigma_i^2 = \text{var}\{H|\omega_i\} \qquad i=1,2$$

با داشتن معیارهای انتخاب (criterion):

$$h(\underline{x}) = \underline{v}^T \underline{x} + v_0$$
$$\eta_i = E\{H|\omega_i\} = \underline{v}^T \underline{m}_i + v_0$$
$$\sigma_i^2 = \text{var}\{H|\omega_i\} = \underline{v}^T [\Sigma]_i \underline{v}$$
$$f(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2) \sim$$ تابع شاخص

پس مسئله پیدا کردن جهت $\underline{v}$ و نیز اعمال تابع f است که می‌تواند تابعهای مختلف باشد ، min probability of error ، Fisher ، ....

این مسئله ، یک مسئله بهینه سازی است:

$$\text{To find } \underline{v} \longrightarrow \frac{\partial f}{\partial \underline{v}} = 0 , \frac{\partial f}{\partial v_0} = 0$$

طبق chain rule داریم:

$$(\text{I})\ \frac{\partial f}{\partial \underline{v}} = \frac{\partial f}{\partial \eta_1} \cdot \frac{\partial \eta_1}{\partial \underline{v}} + \frac{\partial f}{\partial \eta_2} \cdot \frac{\partial \eta_2}{\partial \underline{v}} + \frac{\partial f}{\partial \sigma_1^2} \cdot \frac{\partial \sigma_1^2}{\partial \underline{v}} + \frac{\partial f}{\partial \sigma_2^2} \cdot \frac{\partial \sigma_2^2}{\partial \underline{v}}$$

$$(\text{II})\ \frac{\partial f}{\partial v_0} = \frac{\partial f}{\partial \eta_1} \cdot \frac{\partial \eta_1}{\partial v_0} + \frac{\partial f}{\partial \eta_2} \cdot \frac{\partial \eta_2}{\partial v_0} + \frac{\partial f}{\partial \sigma_1^2} \cdot \frac{\partial \sigma_1^2}{\partial v_0} + \frac{\partial f}{\partial \sigma_2^2} \cdot \frac{\partial \sigma_2^2}{\partial v_0}$$

و داریم:

$$\frac{\partial \eta_i}{\partial \underline{v}} = \underline{m}_i \qquad \frac{\partial \sigma_i^2}{\partial \underline{v}} = 2[\Sigma]_i \underline{v}$$

$$\frac{\partial \eta_i}{\partial v_0} = 1 \qquad \frac{\partial \sigma_i^2}{\partial v_0} = 0$$

با جایگذاری داریم :

$$(I)\quad \boxed{\frac{\partial f}{\partial \eta_1}\underline{m}_1 + \frac{\partial f}{\partial \eta_2}\underline{m}_2 + \frac{\partial f}{\partial \sigma_1^2}(2\Sigma_1 \underline{V}) + \frac{\partial f}{\partial \sigma_2^2}(2\Sigma_2 \underline{V}) = 0}$$

$$(II)\quad \frac{\partial f}{\partial \eta_1} + \frac{\partial f}{\partial \eta_2} = 0 \Longrightarrow \boxed{\frac{\partial f}{\partial \eta_1} = -\frac{\partial f}{\partial \eta_2}}$$

بنابراین جهت بردار $\underline{V}$ اینطور بدست می‌آید:

$$\boxed{\underline{V} = \left[s\,[\Sigma]_1 + (1-s)[\Sigma]_2 \right]^{-1} (\underline{m}_2 - \underline{m}_1)}$$

$$\boxed{\text{where: } s = \frac{\frac{\partial f}{\partial \sigma_1^2}}{\frac{\partial f}{\partial \sigma_1^2} + \frac{\partial f}{\partial \sigma_2^2}}}$$

و مقدار V_0 بهینه اینطور بدست می‌آید:

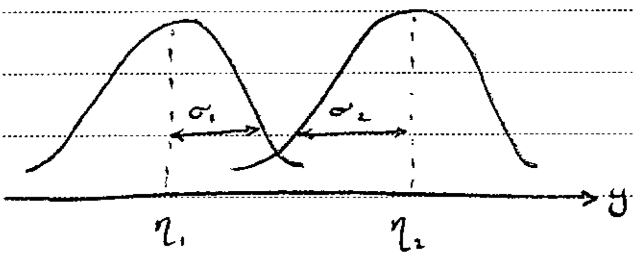
$$V_0 : \frac{\partial f}{\partial \eta_1} + \frac{\partial f}{\partial \eta_2} = 0$$

$$\underbrace{(m_2 - m_1)^T \Sigma^{-1}}_{\underline{V}}\ \underline{x} + \underbrace{\frac{1}{2}(m_1^T \Sigma^{-1} m_1 - m_2^T \Sigma^{-1} m_2) - \ln \frac{P_1}{P_2}}_{V_0} \mathop{\gtrless}_{\omega_1}^{\omega_2} 0$$

$$\Sigma \longrightarrow s[\Sigma]_1 + (1-s)[\Sigma]_2$$

Fisher Criterion .

$$f = \frac{(\eta_1 - \eta_2)^2}{\sigma_1^2 + \sigma_2^2}$$



برای جدا کردن دو توزیع
* فاصله میانگینها زیاد
* فاصله نمونه‌ها از میانگین کم

$y = \underline{v}^T x$

$\sigma_i \downarrow \Rightarrow f \uparrow$

$|\eta_1 - \eta_2| \uparrow \Rightarrow f \uparrow$

محاسبات:

$$s = \frac{\frac{\partial f}{\partial \sigma_1^2}}{\frac{\partial f}{\partial \sigma_1^2} + \frac{\partial f}{\partial \sigma_2^2}}$$

$$\frac{\partial f}{\partial \sigma_1^2} = \frac{\partial f}{\partial \sigma_2^2} = \frac{-(\eta_1 - \eta_2)^2}{(\sigma_1^2 + \sigma_2^2)^2} \equiv a \Longrightarrow s = \frac{a}{a+a} = \frac{1}{2}$$

$$\underline{v} = \left(\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2\right)^{-1}(m_2 - m_1)$$

Fisher Classifier:

$$\begin{cases} h(\underline{x}) = \underline{v}^T \underline{x} = (\underline{m}_2 - \underline{m}_1)^T \left(\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2\right)^{-1} \underline{x} \\ v_0 : \frac{\partial f}{\partial \eta_1} + \frac{\partial f}{\partial \eta_2} = 0 \end{cases}$$

$$f = \frac{P_1 \eta_1^2 + P_2 \eta_2^2}{P_1 \sigma_1^2 + P_2 \sigma_2^2}$$

به نوعی پراکندگی کلاسها را نشان می‌دهد، یعنی این نسبت :

$$\frac{\text{between-class scatter}}{\text{within-class scatter}}$$

محاسبات :

$$\frac{\partial f}{\partial \sigma_i^2} = \frac{-P_i\,(P_1 \eta_1^2 + P_2 \eta_2^2)}{(P_1 \sigma_1^2 + P_2 \sigma_2^2)^2}$$

$$S = \frac{\partial f / \partial \sigma_1^2}{\partial f / \partial \sigma_1^2 + \partial f / \partial \sigma_2^2} = P_1$$

$$\underline{v} = (P_1\,[\Sigma]_1 + \underbrace{(1 - P_1)}_{P_2}\,[\Sigma]_2)^{-1}(\underline{m}_2 - \underline{m}_1)$$

* اگر N عدد داشته باشیم می‌توان برای محاسبه جهت $\underline{v}$ ، مقادیر $[\Sigma]$ را تخمین زد:

$$[\hat{\Sigma}]^i = \frac{1}{N_i - 1}\left(\sum_{e=1}^{N_i} (\underline{x}_e^i - \underline{m}_e^i)(x_e^i - \underline{m}_e^i)^T\right)$$

$$\frac{\partial f}{\partial \eta_i} = \frac{2 P_i \eta_i}{P_1 \sigma_1^2 + P_2 \sigma_2^2}$$

$$\frac{\partial f}{\partial \eta_1} + \frac{\partial f}{\partial \eta_2} = 0$$

$$\Rightarrow \frac{2 P_1 \eta_1 + 2 P_2 \eta_2}{P_1 \sigma_1^2 + P_2 \sigma_2^2} = 0 \xrightarrow{\eta_i = \underline{v}^T m_i + v_0}$$

$$\Rightarrow 2 P_1 (\underline{v}^T m_1 + v_0) + 2 P_2 (\underline{v}^T m_2 + v_0) = 0 \Longrightarrow$$

$$\underline{v}^T (P_1 \underline{m}_1 + P_2 \underline{m}_2) + (P_1 + P_2) v_0 = 0$$

$$P_1 + P_2 = 1$$

$$\Rightarrow v_0 = -\underline{v}^T (P_1 \underline{m}_1 + P_2 \underline{m}_2)$$

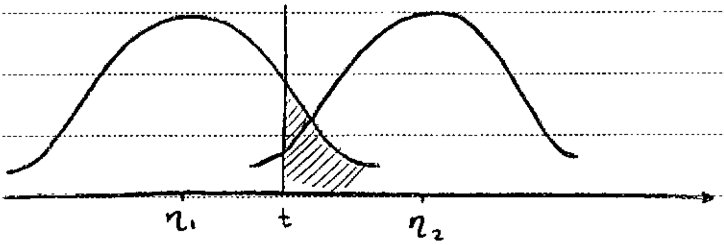
Minimum Probability of Error.

$H(\underline{x}) = \underline{v}^T \underline{X} + v_0$

$n \gg 30 \longrightarrow$ قضیه حد مرکزی $\longrightarrow$ توزیع نرمال

$\eta_i = E\{H \mid \omega_i\} = \underline{v}^T \underline{m}_i + v_0$

$\sigma_i^2 = var\{H \mid \omega_i\} = \underline{v}^T [\Sigma]_i^{-1} \underline{v}$



$$f = \overset{\text{min Prob. of error}}{\varepsilon}$$

$$\varepsilon = P_1 \int_{\frac{-\eta_1}{\sigma_1}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) . d\xi + P_2 \int_{-\infty}^{\frac{\eta_2}{\sigma_2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) . d\xi$$

$$\frac{d}{dx}\left[\int_{g_1(x)}^{g_2(x)} f(t) . dt\right] = f(g_2(x)) . g_2'(x) - f(g_1(x)) . g_1'(x)$$

$$\frac{\partial f}{\partial \sigma_1^2} = \frac{\partial \varepsilon}{\partial \sigma_1^2} = 0 - \frac{P_1}{\sqrt{2\pi}} \exp\left(-\frac{(-\eta_1/\sigma_1)^2}{2}\right) \frac{\partial}{\partial \sigma_1^2}\left(-\frac{\eta_1}{\sigma_1}\right)$$

$$= -\frac{P_1}{\sqrt{2\pi}} . \exp\left(-\frac{\eta_1^2}{2\sigma_1^2}\right) . \underbrace{\frac{\partial}{\partial \sigma_1^2}\left(-\frac{\eta_1}{\sqrt{\sigma_1^2}}\right)}$$

$$+\frac{1}{2}\eta_1 (\sigma_1^2)^{-\frac{3}{2}} \longleftarrow \frac{\partial}{\partial \sigma_1^2}\left(-\eta_1 . (\sigma_1^2)^{-\frac{1}{2}}\right)$$

Subject:

Year. Month. Date. ( )

$$= -\frac{P_1}{2\sqrt{2\pi}\,\sigma_1}\cdot\exp\left(-\frac{\eta_1^2}{2\sigma_1^2}\right)\cdot\left(\frac{\eta_1}{\sigma_1^2}\right)$$

$$\frac{\partial f}{\partial \sigma_2^2} = \frac{\partial \varepsilon}{\partial \sigma_2^2} = +\frac{P_2}{2\sqrt{2\pi}\,\sigma_2}\cdot\exp\left(-\frac{\eta_2^2}{2\sigma_2^2}\right)\cdot\left(\frac{\eta_2}{\sigma_2^2}\right)$$

$$\frac{\partial f}{\partial \eta_1} = \frac{\partial \varepsilon}{\partial \eta_1} = \frac{P_1}{\sqrt{2\pi}\,\sigma_1}\exp\left\{-\frac{\eta_2^2}{2\sigma_1^2}\right\}$$

$$\frac{\partial f}{\partial \eta_2} = \frac{\partial \varepsilon}{\partial \eta_2} = -\frac{P_2}{\sqrt{2\pi}\,\sigma_2}\cdot\exp\left\{-\frac{\eta_2^2}{2\sigma_2^2}\right\}$$

$$\frac{\partial f}{\partial \eta_1} + \frac{\partial f}{\partial \eta_2} = 0 \Rightarrow \frac{P_1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\frac{\eta_1^2}{2\sigma_1^2}\right) = \frac{P_2}{\sqrt{2\pi}\,\sigma_2}\exp\left(-\frac{\eta_2^2}{2\sigma_2^2}\right)$$

طبق این فرمول، v_0 نقطه جایی است که $\varepsilon_1 = \varepsilon_2$

$$S = \frac{\dfrac{\partial f}{\partial \sigma_1^2}}{\dfrac{\partial f}{\partial \sigma_1^2} + \dfrac{\partial f}{\partial \sigma_2^2}} = \frac{\cancel{\dfrac{P_1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\dfrac{\eta_1^2}{2\sigma_1^2}\right)}\left(\dfrac{-\eta_1}{\sigma_1^2}\right)}{\underbrace{\cancel{\dfrac{P_1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\dfrac{\eta_1}{2\sigma_1^2}\right)}}\left(\dfrac{-\eta_1}{\sigma_1^2} + \dfrac{\eta_2}{\sigma_2^2}\right)}$$

قابل فاکتورگیری ←

$$S = \frac{-\eta_1/\sigma_1^2}{-\eta_1/\sigma_1^2 + \eta_2/\sigma_2^2} = \frac{1}{1 + \left(\dfrac{\eta_2/\sigma_2^2}{-\eta_1/\sigma_1^2}\right)}$$

$$\eta_i = \underline{v}^T \underline{m}_i + v_0 \underset{\omega_1}{\overset{\omega_2}{\gtrless}} 0 \Longrightarrow \eta_1 < 0 \,,\quad \eta_2 > 0$$

$$S = \frac{1}{1+\text{positive}} \Rightarrow 0 < S < 1$$

PAPCO

پس ابتدا باید η_1، η_2، σ_1^2 و σ_2^2 به S و بدست آوردن v_0 داریم:

$$v_0 = \frac{s\,\sigma_1^2\,\underline{v}^T\underline{m}_2 + (1-s)\,\sigma_2^2\,\underline{v}^T\underline{m}_1}{s\,\sigma_1^2 + (1-s)\,\sigma_2^2}$$

با توجه به اینکه مقدار S محدودیت دارد در بازهٔ $0 < S < 1$ محدود است، برای آن یک راه حل Iterative بدست می‌آید:

1. choose S

$$\underline{v} = (s\,\Sigma_1 + (1-s)\,\Sigma_2)^{-1}(\underline{m}_2 - \underline{m}_1)$$

2. using $\underline{v}$

$$\sigma_i^2 = \underline{v}^T\Sigma_i\,\underline{v}$$

v_0

$$\eta_i = \underline{v}^T\underline{m}_i + v_0$$

3. calculate ε

$$\varepsilon = P_1 \int_{-\frac{\eta_1}{\sigma_1}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}\, d\xi + P_2 \int_{-\infty}^{-\frac{\eta_2}{\sigma_2}} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}\, d\xi$$

4. repeat for new S

همچنین S بدست آمده که کمترین خطا را بدهد، جواب است.

اگر با روش شمارش خطا بخواهیم این روال را پیاده‌سازی کنیم:

$$N \text{ sample} : \begin{cases} N_1 \in \omega_1 \\ N_2 \in \omega_2 \end{cases}$$

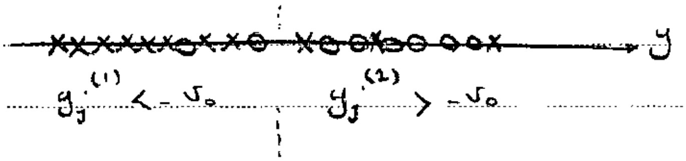
1. compute $\hat{\underline{M}}_1$, $\hat{\underline{M}}_2$, $\hat{\Sigma}_1$, $\hat{\Sigma}_2$

2. for a given S $(0 < S < 1)$

$$\underline{v} = (S\,\hat{\Sigma}_1 + (1-S)\,\hat{\Sigma}_2)^{-1}\,(\underline{m}_2 - \underline{m}_1)$$

3. $\underline{v}^T \underline{x} + v_0 \xrightarrow{\text{linear}} y + v_0 \Rightarrow y_j^{(i)} = \underline{v}^T x_j^{(i)}$ $\quad i=1,2$
$j=1,\dots,N_i$

4.



$y_j^{(1)} < -v_0$ $\quad$ $y_j^{(2)} > -v_0$

Choose the point where error is smallest

5. change S

این روش از نمونه‌ها هم در یادگیری و هم در آزمایش استفاده می‌کند.

در روش سوم ، نمونه‌ها را به دو دسته تقسیم می‌کنیم ، تعدادی از آنها را برای یادگیری استفاده می‌کند
و ما بقی برای آزمایش استفاده می‌کند.

Sample Set : $\begin{cases} \text{Training Set} \\ \text{Test Set} \end{cases}$

خطا ← روش II ← تمام داده‌ها را دید ← کمترین خطا
Biased

روش III ← آزمایش توسط داده‌های جدید ← خطای بیشتر
نزدیک به مقدار واقعی

* طراحی تمایز دهنده‌ی خطی

• سنجش MSE

Mean Square Error

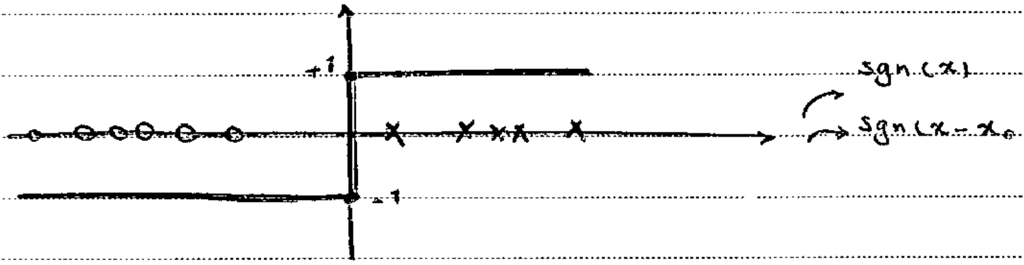
•

$\underline{h}(\underline{x}) \longrightarrow$ سطای

$h(\underline{x}) = \underline{v}^T \underline{x} + v_0$

$\gamma(\underline{x}) \longrightarrow$ desired form of our classifier

$$\gamma(\underline{x}) = \begin{cases} -1 , & \underline{x} \in \omega_1 \\ +1 , & \underline{x} \in \omega_2 \end{cases}$$



•

$$\bar{\varepsilon}^2 = E\left\{ (H(\underline{x}) - \gamma(\underline{x}))^2 \right\}$$

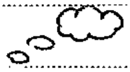
$$= E\left\{ H^2(\underline{x}) - 2\gamma(\underline{x}).H(\underline{x}) + \gamma^2(\underline{x}) \right\}$$

$$= \underbrace{E\left\{ H^2(\underline{x}) \right\}}_{II} - 2 \underbrace{E\left\{ \gamma(\underline{x}) H(\underline{x}) \right\}}_{I} + \underbrace{\gamma^2(\underline{x})}_{\text{deterministic}}$$

$$\gamma(\underline{x}) = \begin{cases} -1 \; , & \underline{x} \in \omega_1 \\ +1 \; , & \underline{x} \in \omega_2 \end{cases}$$

I) $E\{H(\underline{x})\,\gamma(\underline{x})\} = P_1 . E\{H(\underline{x})\,(-1)\,|\,\omega_1\} + P_2 . E\{H(\underline{x})\,(+1)\,|\,\omega_2\}$

$= -P_1 . \eta_1 + P_2\,\eta_2$



$E\{X\} = m$

$E\{(X-m)^2\} = \operatorname{var}\{X\}$

$E\{X^2\} = \sigma_X^2 + \eta_X^2$

II) $E\{H^2(\underline{x})\} = P_1\,(\sigma_1^2 + \eta_1^2) + P_2\,(\sigma_2^2 + \eta_2^2)$

$\bar{\varepsilon}^2 = \left(P_1\,(\sigma_1^2+\eta_1^2) + P_2\,(\sigma_2^2+\eta_2^2)\right) - 2\,(-P_1\eta_1 + P_2\eta_2)$

$\underline{V} = (s\,\Sigma_1 + (1-s)\,\Sigma_2)^{-1}\,(\underline{m}_2 - \underline{m}_1)$

$$s = \frac{\dfrac{\partial f}{\partial \sigma_1^2}}{\dfrac{\partial f}{\partial \sigma_1^2} + \dfrac{\partial f}{\partial \sigma_2^2}} \quad \Big| \quad \Longrightarrow s = \frac{P_1}{P_1 + P_2} = P_1$$

$$\frac{\partial f}{\partial \sigma_i^2} = P_i$$

$$V_0 : \frac{\partial f}{\partial \eta_1} + \frac{\partial f}{\partial \eta_2} = 0 \longrightarrow \ldots$$

$\gamma(\underline{x}) = q_2(\underline{x}) - q_1(\underline{x})$ post priori probability

$$I)\quad E\{H(\underline{x})\,\gamma(\underline{x})\} = \int_{-\infty}^{+\infty} h(\underline{x}) \times \frac{P_2\,p_2(\underline{x}) - P_1\,p_1(\underline{x})}{\cancel{p(\underline{x})}}\,\cancel{p(\underline{x})}\,d\underline{x}$$

$$= P_2\,\eta_2 - P_1\,\eta_1$$

مانند حالت قبل شد...

Subject:

Year. Month. Date. ( )

• اگر بتوان $[u][u]^T$ را به $[I]$ تبدیل کرد، کار آسان می‌شود. برای آن
باید میانگین را به صفر برد که اینها با یک تبدیل ممکن است.

$$\left.\begin{aligned} &\frac{1}{N}\sum_{i=0}^{N} \underline{y}_i = 0 \\ &\frac{1}{N}\sum_{i=0}^{N} \underline{y}_i \underline{y}_i^T = [I] \end{aligned}\right\} \Leftarrow \underline{y}_j = [a]^T \underline{x}_j + \underline{b}$$

$$\left.\begin{aligned} &\underline{y}_j = [a]^T \underline{x}_j + \underline{b} \\ &\underline{d}_i = [a]^T \underline{m}_i + \underline{b} \\ &[K]_i = [a][\Sigma]_i[a]^T \end{aligned}\right\} \Rightarrow [u][u]^T = N[I] \Rightarrow \underline{w} = \frac{1}{N}[u']\underline{\Gamma}'$$

$$\underline{w} = \frac{1}{N}[u']\underline{\Gamma}'$$

$$[u'] = [\underline{y}_1, \ldots, \underline{y}_N] = [\underline{Z}_1', \ldots, \underline{Z}_N']$$

$$\underline{w} = \begin{bmatrix} v_0 \\ \underline{v} \end{bmatrix}$$

$$w_0 = \frac{1}{N}\left[\sum_{j=1}^{N_2} \gamma(\underline{Z}_j') - \sum_{j=N_2+1}^{N} \gamma(\underline{Z}_j')\right]$$

$$[w_1' \ldots w_n']^T = \frac{1}{N}\left[\sum_{j=1}^{N_2} \gamma(y_j)\,\underline{y}_j - \sum_{j=N_2+1}^{N} \gamma(y_j)\,\underline{y}_j\right]$$

• می‌خواهیم این رابطه برقرار باشد:

$$h(Z) = \underline{w}^T \underline{Z} \geq 0$$

پس فرض می‌کنیم که:

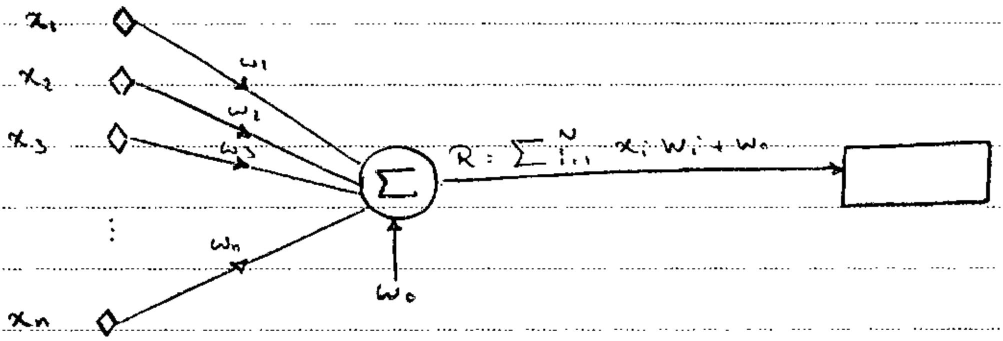
$$\gamma(\underline{Z}) = 1$$

با قرار دادن در رابطه‌ی قبل داریم:

$$W_0 = \frac{N_2}{N} - \frac{N_1}{N} = P_2 - P_1$$

$$\underline{W} = P_2\,\underline{d}_2 - P_1\,\underline{d}_1$$

• گاهی اوقات مایلیم $\underline{W}$ را به صورت Iterative پیدا کنیم. این روش بسیار نزدیک به Perceptron Approach می‌باشد.



یکی از روش‌های وزن دهی به نام Reward – Punishment معروف است.

N_1 sample $\in x_1$

N_2 sample $\in x_2$

$N_1 + N_2 = N$

$$h(\underline{x}) = \underline{W}^T \underline{x} + \underline{W_0} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$W(1)$ to be initial vector

<u>K^{th} step:</u>

I.) $\underline{x}(k) \in \omega_1$ & $\underline{x}^T(k)\,\underline{W}(k) \leqslant 0$

replace $W(k)$ by $\underline{W}(k+1) = \underline{W}(k) + C\,\underline{x}(k)$

$C \equiv$ correction increment

II) $x(k) \in \omega_2$ & $\underline{x}^T(k)\,\underline{w}(k) \geq 0$

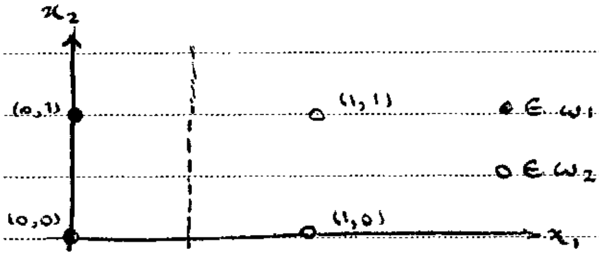
$$\underline{W}(k+1) = \underline{W}(k) - C \cdot \underline{x}(k)$$

III Otherwise

$$\underline{W}(k+1) = \underline{W}(k)$$

ثابت شده است که اگر دو کلاس به صورت خطی جداپذیر باشند، این الگوریتم در مقدار $\underline{W}$ همگراست.

☑



$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \omega_1$

$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \omega_2$

$h(\underline{x}) = \underline{w}^T \underline{z} = \underline{v}^T \underline{x} + v_0$

$\omega_1 : \{ (0\ 0\ 1)^T, (0\ 1\ 1)^T \}$

$\omega_2 : \{ (1\ 0\ 1)^T, (1\ 1\ 1)^T \}$

$\underline{w}(1) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ $\quad C = 1$

* $\underline{x}(1) \in \omega_1$: $\underline{x}(1) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $w^T(1)\,\underline{x}(1) = 0$ $\Rightarrow \underline{w}(2) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

* $\underline{x}(2) \in \omega_1$: $\underline{x}(2) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\underline{w}^T(2)\,\underline{x}(2) = 1$ $\Rightarrow \underline{W}(3) = \underline{W}(2) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

* $\underline{x}(3) \in \omega_2$ $\quad \underline{x}(3) = \begin{bmatrix}1\\0\\1\end{bmatrix}$, $\underline{W}^T(3)\,\underline{x}(3) = 1$ $\Longrightarrow W(4) = \begin{bmatrix}0\\0\\1\end{bmatrix} - 1\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}-1\\0\\0\end{bmatrix}$

* $\underline{x}(4) \in \omega_2$ $\quad \underline{x}(4) = \begin{bmatrix}1\\1\\1\end{bmatrix}$, $\underline{w}^T(4)\,\underline{x}(4) = -1$ $\Longrightarrow W(5) = W(4)$

* $\underline{x}(5) \in \omega_1$ $\quad \underline{x}(5) = \begin{bmatrix}0\\0\\1\end{bmatrix}$, $\underline{W}^T(5)\,\underline{x}(5) = 0$ $\Longrightarrow W(6) = \begin{bmatrix}-1\\0\\0\end{bmatrix} + 1\begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}-1\\0\\1\end{bmatrix}$

* $\underline{x}(6) \in \omega_1$ $\quad \underline{x}(6) = \begin{bmatrix}0\\1\\1\end{bmatrix}$, $\underline{W}^T(6)\,\underline{x}(6) = 1$ $\Longrightarrow W(7) = W(6)$

* ...

$\underline{W}(12) = \begin{bmatrix}-2\\0\\1\end{bmatrix}$

$\underline{x} \in \omega_2$: $(0 \; 0 \; 1)\,\underline{W}(12) = 1 > 0$ ✓

$(0 \; 1 \; 1)\,\underline{W}(12) = 1 > 0$ ✓

$\underline{x} \in \omega_1$: $(1 \; 0 \; 1)\,\underline{W}(12) = -1 < 0$ ✓

$(1 \; 1 \; 1)\,\underline{W}(12) = -1 < 0$ ✓

* $\underline{W} = \underline{W}(12) = [-2 \quad 0 \quad 1]^T$

$h(\underline{x}) = \underline{x}^T\underline{w} = (x_1 \quad x_2 \quad 1)\begin{pmatrix}-2\\0\\1\end{pmatrix} = -2x_1 + 1$

$h(\underline{x}) = 0 \longrightarrow \boxed{x_1 = \frac{1}{2}}$

• حال اگر این قانون را طور دیگری نمایش دهیم:

$$h(\underline{x}) = \underline{v}^T \underline{x} + v_0 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$$= \underline{W}^T x > 0$$

if $\underline{x} \in \omega_2$ then $x = \begin{bmatrix} -x_1 \\ \vdots \end{bmatrix}$

$$\underline{W}(k+1) = \begin{cases} \underline{W}(k) & , \ \underline{x}^T(k)\,\underline{W}(k) > 0 \\ \\ \underline{W}(k) + c\,\underline{x}(k) & , \ \underline{x}^T(k)\,\underline{W}(k) \leqslant 0 \end{cases}$$

• نکاتی در رابطه با الگوریتم Perceptron

* Fixed Increment Alg.

$c \equiv$ constant

* Absolute Correction Alg.

می‌خواهیم مقدار c را طوری تنظیم کنیم تا شرط خاصی را ارضا کند.

$\underline{k^{th} \text{ step}}$

$$\underline{W}^T(k)\ \underline{x}(k) \leqslant 0$$

$$\underline{W}^T(k+1)\ \underline{x}(k) = \left(\underline{W}(k) + c\,\underline{x}(k)\right)^T \underline{x}(k) > 0$$

$$\Rightarrow c(k) > \frac{|\underline{W}^T(k) \cdot x(k)|}{\underline{x}^T(k) \cdot \underline{x}(k)}$$

* Fractional Correction Alg.

$$|\underline{W}^T(k) \cdot x(k) - \underline{W}^T(k+1) \cdot x(k)| = \lambda\, |\underline{W}^T(k)\, x(k)|$$

$$\Rightarrow c = \lambda \frac{|\underline{W}^T(k)\ \underline{x}(k)|}{\underline{x}^T(k) \cdot \underline{x}(k)}$$

• ثابت شده است که اگر داده‌ها به طور خطی جداپذیر باشند، الگوریتم Perceptron همگرا می‌شود.

اگر الگوریتم مدتی طولانی مشغول بود، یا جواب به سختی به دست آمده است یا داده‌ها به طور خطی جداپذیر نیست.

برای داشتن ساختار مناسب راه‌های مختلفی وجود دارد:

- میانگین W‌ها را نگه داریم تا بلکه یک جواب تقریباً مناسب موجود باشد.
- غیرخطی بودن به نحوی مستقل داده شود.
- هرجایی که الگوریتم قطع شود، جواب بهترین تا آن لحظه داده شود.

Packet Algorithm ←

• Packet Algorithm

$\underline{W}_s$ ≡ Sorted Weight Vector (Packet)

h_s ≡ History Counter

الگوریتم:

1. initialize $\underline{W}(0)$, $h_s = 0$
2. @ the k^{th} Step
   - compute & update the weight vector $\underline{W}(k+1)$ (based on the perceptron rule)
   - use $\underline{W}(k+1)$ to test # of training samples classified correctly are counted $= h$
   - if $h > h_s$ then
     $\underline{W}_s = \underline{W}(k+1)$
     $h_s = h$
   - continue the iteration

- گاهی اوقات تغییر وزن براساس شاخصی تعیین می‌شود که به آن Gradient می‌گویند.

* The Gradient Technique

-
$$\nabla_{\underline{y}} f(\underline{y}) = grad\,[f(\underline{y})] = \begin{bmatrix} \frac{df}{dy_1} \\ \frac{df}{dy_2} \\ \cdots \\ \frac{df}{dy_n} \end{bmatrix}$$

- جهت طی بردار به سمت بیشترین نرخ تغییرات خواهد بود. ← Positive Gradient

- برای اعمال به PR می‌توان یک criterion ترتیب کرد و از Negative Grad استفاده کرد تا تغییرات به سمت مینیمم کردن پیش برود و به آن روشهای Gradient Descend می‌گویند.

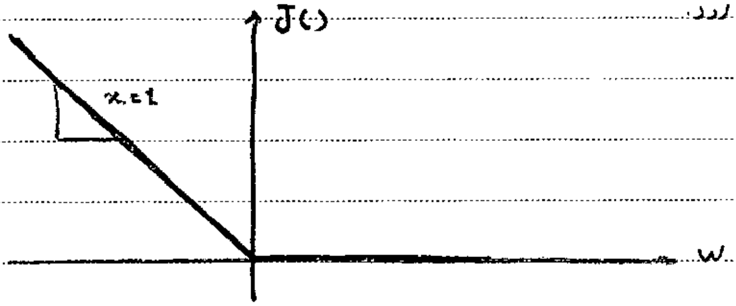
-
$$h(\underline{x}) = \underline{w}^T \underline{x} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$$= \underline{w}^T \underline{x} > 0$$

criteria function

$$J(\underline{w}, \underline{x}) = (|\underline{w}^T \underline{x}| - \underline{w}^T \underline{x})$$

solution در جایی است که مقدار این تابع صفر شود. برای پیدا کردن جواب باید تغییرات در راستای کمینه کردن J پیش رود.



$$\underline{W}(k+1) = \underline{W}(k) + C(k).\ \underline{x}(k)$$

$$\underline{W}(k+1) = \underline{W}(k) - c\ \nabla_{\underline{w}} J(\underline{w}, \underline{x})$$

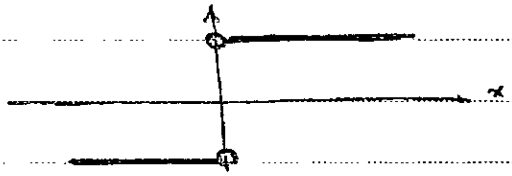
• مثل Perceptron

$$J(\underline{w}, \underline{x}) = \frac{1}{2}\left(|\underline{w}\underline{x}| - \underline{w}\underline{x}\right)$$

$$\underline{W}(k+1) = \begin{cases} \underline{W}(k) & , \underline{x}^T(k)\ \underline{W}(k) > 0 \\ \underline{W}(k) + c\,\underline{x}(k) & , \underline{x}^T(k)\ \underline{W}(k) \leqslant 0 \end{cases}$$

$$\frac{\partial J(\underline{w}, \underline{x})}{\partial \underline{w}} = \frac{1}{2}\left(\underline{x}.\ \mathrm{sgn}\ (\underline{w}^T\underline{x}) - \underline{x}\right)$$

$$\mathrm{sgn}(x) = \begin{cases} +1 & , x > 0 \\ -1 & , x < 0 \end{cases}$$



$$\underline{W}(k+1) = \underline{W}(k) + \frac{c}{2}\left\{\underline{x}(k) - \underline{x}(k).\ \mathrm{sgn}\ (\underline{W}^T(k).\ \underline{x}(k))\right\}$$

$$= \underline{W}(k) + c \times \begin{cases} 0 & , \underline{W}^T(k).\ \underline{x}(k) > 0 \\ \underline{x}(k) & , \underline{W}^T(k).\ \underline{x}(k) < 0 \end{cases}$$

مانند روش Reward - Punishment عمل می‌کند.

• criteria function 2

$$J(\underline{w}, \underline{x}) = \frac{1}{4\underline{x}^T\underline{x}}\left(|\underline{w}^T\underline{x}|^2 - |\underline{w}^T\underline{x}|\ \underline{w}^T\underline{x}\right)$$

$$\frac{\partial J(\sim)}{\partial \underline{w}} = \frac{1}{4\underline{x}^T\underline{x}}\left[2\,|\underline{w}^T\underline{x}|\,\underline{x}\cdot \operatorname{sgn}(\underline{w}^T\underline{x}) - |\underline{w}^T\underline{x}|\,\underline{x} - (\underline{w}^T\underline{x})\,\underline{x}\,\operatorname{sgn}(\underline{w}^T\underline{x})\right]$$

$$= \frac{1}{2\underline{x}^T\underline{x}}\left[\,|\underline{w}^T\underline{x}|\,\underline{x}\ \operatorname{sgn}(\underline{w}^T\underline{x}) - |\underline{w}^T\underline{x}|\,\underline{x}\,\right]$$

تقریباً معادل Absolute Correction Alg. می باشد.

• روش MSE

$$h(\underline{x}) = \underline{v}^T\underline{x} + v_0 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$$= \underline{w}^T\underline{\underline{z}} > 0$$

$$= \sum_{i=0}^{n} w_i z_i > 0$$

$$\bar{\varepsilon}^2 = E\{(h(\underline{x}) - \gamma(\eta))^2\}$$

sample های مرزی خیلی مهم‌تر است.
روی MSE تأثیر بیشتری باید داشته باشند، درحالی که در سایر روش‌ها به sample های مرکزی اهمیت بیشتری داده می‌شود.
این روش سعی در اصلاح این وضع دارد:

*

$$\bar{\varepsilon}^2 = \frac{1}{N}\sum_{j=1}^{N}\left\{\underline{w}^T\underline{z}_j - |\underline{w}^T\underline{z}_j|\right\}^2$$

$$\underline{w}^T\underline{z}_j \begin{cases} >0 & , \bar{\varepsilon}^2 \sim cte \longrightarrow \text{دسته‌بندی درست} \\ <0 & , \bar{\varepsilon}^2 \uparrow \longrightarrow \text{دسته‌بندی خطا} \end{cases}$$

*

$$\bar{\varepsilon}^2 = \frac{1}{N}\sum_{j=1}^{N}\left\{\operatorname{Sgn}(\underline{w}^T\underline{z}_j) - 1\right\}^2$$

شمارش تعداد خطا

$$\bar{\varepsilon}^2 = E\{(h(\underline{x}) - \gamma(\underline{x}))^2\}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\{\underline{w}^T \underline{x}_j - \gamma(\underline{x}_j)\}^2$$

$$= \frac{1}{N}([u]^T\underline{w} - \underline{I})([u]^T\underline{w} - \underline{I})$$

$$[u] = [\underline{x}_1 \; \dots \; \underline{x}_N]$$
$$\underline{I} = [\gamma(\underline{x}_1) \; \dots \; \gamma(\underline{x}_N)]^T$$

$$\frac{\partial \bar{\varepsilon}^2}{\partial \underline{w}} = \frac{2}{N}[u]([u]^T\underline{w} - \underline{I})$$

$$\underline{w}(k+1) = \underline{w}(k) - c\,\nabla_{\underline{w}} J(\underline{w}, \underline{x})$$

$$\underline{w}(k+1) = \underline{w}(k) - \frac{2c}{N}\{[u]^T\underline{w}(k) - |[u]^T\underline{w}(k)|\}$$

- حال اگر بخواهیم یک criterion داشته باشیم که بتوان از آن مشتق بگیریم از روش نیوتن استفاده می‌کنیم.

Newton Method

$$\underline{w}(k+1) = \underline{w}(k) - [H]^{-1}\nabla_{\underline{w}} J(\underline{w})$$

$$[H] \equiv \text{Hessian Matrix} \sim \frac{\partial J(\sim)}{\partial w_i . \partial w_j}$$

* الگوریتم Ho - Kashyap

sample : $\underline{x}_1 , \dots , \underline{x}_N$

$x_i \in \omega_1 \qquad i = 1, \dots, N_1$

$x_i \in \omega_2 \qquad i = N_1+1 , \dots , N$

$$h(\underline{x}) = \underline{V}^T \underline{x} + V_0 \quad \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

$$= \underline{W}^T \underline{x} > 0$$

می‌خواهیم این مقدار برابر « b » باشد

حل :

$$\begin{cases} h(\underline{x}) = \underline{W}^T \underline{x} = b \\ \underline{W}^T \underline{x}_i = b_i > 0 \end{cases}$$

$$[u] = [x] = [\underline{x}_1 \quad \underline{x}_2 \quad \cdots \quad \underline{x}_N]$$

$$[x]^T \underline{w} = \underline{b}$$

$$\underline{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} > \underline{0}$$

$$J(\underline{w}, \underline{x}, \underline{b}) = \frac{1}{2} \sum_{j=1}^{N} (\underline{w}^T \underline{x}_j - b_j)^2$$

$$= \frac{1}{2} \| [x]^T \underline{w} - \underline{b} \|^2$$

$$\underline{W}(k+1) = \underline{W}(k) - c \left. \frac{\partial J(\sim)}{\partial \underline{W}} \right|_{\underline{W} = \underline{W}(k)}$$

$$\frac{\partial J}{\partial \underline{w}} = [x] \left([x]^T \underline{w} - \underline{b} \right)$$

$$\frac{\partial J}{\partial b} = -\left([x]^T \underline{w} - \underline{b} \right)$$

- بهینه‌سازی

$$\frac{\partial J}{\partial \underline{w}} = \underline{0} \longrightarrow \underline{w} = ([x][x]^T)^{-1}[x]\,\underline{b} = [x]^{\#}\underline{b}$$

$[x]^{\#} = ([x][x]^T)^{-1}[x]$ ⟶ generalized inverse of $[x]$

$$\underline{b} > 0$$

$$b(k+1) = b(k) - \nabla_{\underline{b}} J(\sim)$$

$$\delta b_i(k) = \begin{cases} 0 & , \left[[x]^T \underline{w} - \underline{b}(k) \right]_i \leqslant 0 \\ 2c\left[[x]^T \cdot \underline{w}(k) - \underline{b}(k) \right] & , \quad \quad > 0 \end{cases}$$

$c = cte > 0$

$$\delta(\underline{b}) = c\left[\left([x]^T \underline{w}(k) - \underline{b}(k) \right) + \left| [x]^T \underline{w}(k) - \underline{b}(k) \right| \right]$$

b تنها زمانی تغییر می‌کند که solution نرسیده باشیم.

$$\delta(\underline{b}) = c\left(\underline{e}(k) - |\underline{e}(k)| \right)$$

-

$$\underline{W} = [x]^{\#} \cdot \underline{b}$$

$$\underline{w}(k+1) = [x]^{\#} \cdot \underline{b}(k+1)$$

$$= [x]^{\#} \cdot \left(\underline{b}(k) + c \nabla_{\underline{b}} J(\sim) \right)$$

$$= [x]^{\#} \cdot \underline{b} + c [x]^{\#} \nabla_{\underline{b}} J(\sim)$$

$$= w(k) + c [x]^{\#} \nabla_{\underline{b}} J(\sim)$$

Ho - Kashyap Algorithm •

1. initialization

$$\underline{b}(1) > \underline{0}$$
$$\underline{w}(1) = [x]^{\#}\, \underline{b}(1)$$

2. K^{th} step

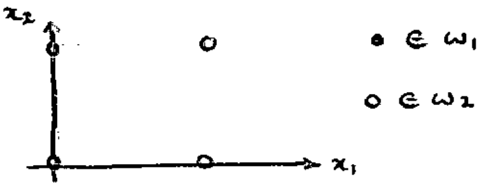
$$\underline{e}(k) = [x]^{T}\underline{w}(k) - \underline{b}(k)$$
$$\underline{w}(k+1) = \underline{w}(k) + c\,[x]^{\#}\,(\underline{e}(k) + |\underline{e}(k)|)$$
$$\underline{b}(k+1) = \underline{b}(k) + c\,(\underline{e}(k) + |\underline{e}(k)|)$$

• اگر مسئله خطی جدا پذیر باشد ← به $\underline{w}$ همگرا می‌شود
اگر مسئله خطی جدا پذیر نباشد ← اگر کلیه عناصر بردار خطا در طی تکرار مثبت باقی
بماند نشان می دهد که مسئله غیرخطی جداپذیر است.

2- class ☑

ω_1: $\left\{ \binom{0}{0}, \binom{0}{1} \right\}$

ω_2: $\left\{ \binom{1}{0}, \binom{1}{1} \right\}$



$\underline{w} = \binom{v_0}{\underline{v}}$ یا $\binom{\underline{v}}{v_0}$

$\underline{x} = \binom{1}{\underline{x}}$ یا $\binom{\underline{x}}{1}$

$$[x] = \begin{bmatrix} 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}_{3\times4} \Rightarrow [x]^{\#} = \underbrace{\left(\underbrace{[x]}_{3\times4}\underbrace{[x]^{T}}_{4\times3}\right)^{-1}}_{3\times3}\underbrace{[x]}_{3\times4} = \frac{1}{2}\begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}_{3\times4}$$

$\in\omega_1$ $\in\omega_2$

$$\underline{b}(1) > \underline{0}$$
$$\underline{b}(1) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad c = 1$$

$$\underline{w}(1) : \underset{3\times 4}{[x]^{\#}}\ \underset{4\times 1}{\underline{b}[1]} : \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

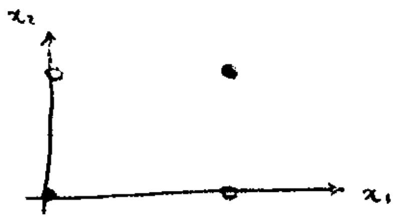
$$[x]^T\ \underline{w}(1) : \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\underline{e}(1) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Longrightarrow \underline{w}(1) \equiv \text{solution} \Rightarrow h(\underline{x}) = -2x_1 + 1$$

XOR ☑

$$\omega_1 : \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

$$\omega_2 : \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$



$$\underline{b}(1) : \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$[x] : \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 \end{bmatrix} \Longrightarrow [x]^{\#} : \frac{1}{2}\begin{bmatrix} -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ \frac{3}{2} & \frac{-1}{2} & \frac{-1}{2} & \frac{-1}{2} \end{bmatrix}$$

$$\underline{w}(1) : [x]^{\#}\ \underline{b}(1) : \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\underline{e}(1) : [x]^T \underline{w}(1) - \underline{b}(1) : \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

چون تمام مولفه های $\bar{e}(1)$ منفی است ⇐ مسئله خطی جداپذیر نیست.

- با استفاده از نمونه‌هایی که به غلط طبقه بندی شده‌اند:

$\underline{w}^T \underline{x} = b \quad , b > 0$

$$J = \frac{1}{2} \sum_{\underline{x} \in Y} \frac{(\underline{w}^T \underline{x} - \underline{b})^2}{\| \underline{x} \|}$$

Y is set of samples for which $\underline{w}^T \underline{x} < b$

$$\frac{\partial J}{\partial \underline{w}} \equiv \nabla_{\underline{w}} J = \sum_{\underline{x} \in Y} \frac{\underline{w}^T \underline{x} - b}{\| \underline{x} \|} \underline{x}$$

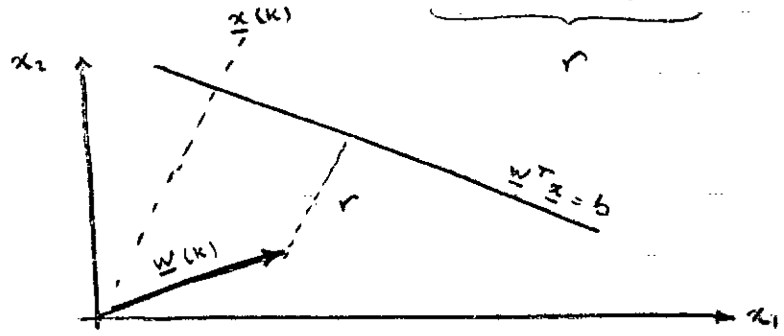
updating rule:

$$\underline{w}(k+1) = \underline{w}(k) - c \nabla_{\underline{w}} J$$

$$\underline{w}(k+1) = \underline{w}(k) - c \sum_{\underline{x} \in Y} \frac{b - \underline{w}^T \underline{x}}{\| \underline{x} \|^2} \underline{x}$$

اگر بخواهیم تک تک داده‌ها استفاده کرد:

$$\underline{w}(k+1) = \underline{w}(k) - c \times \underbrace{\frac{b - \underline{w}^T(k) . x(k)}{\| x(k) \|^2}}_{r} . x(k)$$



ص 226
کتاب Duda & Heart

به تغییر W با هر sample روش Single_Sample Relaxation with Margin
می‌گویند و به کل فرآیند با چند نمونه، روش Batch Relaxation with Margin
اطلاق می‌شود.

* تعمیم روی مسائل چندکلاسه

M-class Problem

$d_i(\underline{x}) = \underline{w}_i^T \, \underline{x}$ i^{th} class

$x \in \omega_i$

$d_i(\underline{x}) > d_j(\underline{x}) \quad \forall j \neq i$

$x(k) \in \omega_i$ k^{th} sample

$d_i(\underline{x}(k)) > d_j(\underline{x}(k)) \quad \forall j \neq i = 1, \dots, M$

$\underline{x}(k) \in \omega_i$

$d_i(\underline{x}) = \underline{w}_i^T(k) \, \underline{x}(k)$

$d_i(\underline{x}(k)) < d_\ell(\underline{x}(k)) \implies \ell^{th}$ classifier is not correct

Learning Rule:

$\underline{W}_i(k+1) = \underline{W}_i(k) + c \, \underline{x}(k)$ → تشویق می‌شود

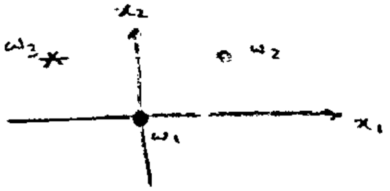
$\underline{W}_\ell(k+1) = \underline{W}_\ell(k) - c \cdot \underline{x}(k)$ → جریمه می‌شود

$\underline{W}_j(k+1) = \underline{W}_j(k)$ $j \neq i, \ell$ — بی تغییر می‌ماند

$w_1 : \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$

$\omega_2 : \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$

$\omega_3 : \left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$



$\underline{w}_1(1) = \underline{w}_2(1) = \underline{w}_3(1) = \begin{vmatrix} 0 \\ 0 \\ 0 \end{vmatrix}$

$c = 1$

$d_i(\underline{x}) = w_i^T \underline{x}$ $i = 1, 2, 3$

* $\underline{x}(1) = \begin{pmatrix}0\\0\\1\end{pmatrix} \in \omega_1$

$$d_1(\underline{x}(1)) = d_2(\underline{x}(1)) = d_3(\underline{x}(1)) = 0$$

$$\text{ideal}: \begin{cases} d_1(\underline{x}(1)) > d_2(\underline{x}(1)) \longrightarrow \times \\ d_1(\underline{x}(1)) > d_3(\underline{x}(1)) \longrightarrow \times \end{cases}$$

$$\text{update}: \begin{cases} \underline{W}_1(2) = \underline{W}_1(1) + c\,\underline{x}(1) = \begin{pmatrix}0\\0\\1\end{pmatrix} \\ \underline{W}_2(2) = \underline{W}_2(1) - c\,\underline{x}(1) = \begin{pmatrix}0\\0\\-1\end{pmatrix} \\ \underline{W}_3(2) = \underline{W}_3(1) - c\,\underline{x}(1) = \begin{pmatrix}0\\0\\-1\end{pmatrix} \end{cases}$$

* $\underline{x}(2) = \begin{pmatrix}1\\1\\1\end{pmatrix} \in \omega_2$

$$\underline{W}_1^T(2)\,.\,\underline{x}(2) = 1$$
$$\underline{W}_2^T(2)\,.\,\underline{x}(2) = -1$$
$$\underline{W}_3(2)\,.\,\underline{x}(2) = -1$$

$$\text{update}: \begin{cases} \underline{W}_1(3) = W_1(2) - c\,.\,\underline{x}(2) = \begin{pmatrix}-1\\-1\\0\end{pmatrix} \\ \underline{W}_2(3) = W_2(2) + c\,.\,\underline{x}(2) = \begin{pmatrix}1\\1\\0\end{pmatrix} \\ \underline{W}_3(3) = W_3(2) - c\,.\,\underline{x}(2) = \begin{pmatrix}-1\\-1\\2\end{pmatrix} \end{cases}$$

* ....

$$\underline{W}_1(8) = \underline{W}_1(7) = \begin{pmatrix}0\\-2\\0\end{pmatrix} \longrightarrow d_1(\underline{x}) = -2x_2$$
$$\underline{W}_2(8) = \underline{W}_2(7) = \begin{pmatrix}2\\0\\-2\end{pmatrix} \longrightarrow d_2(\underline{x}) = 2x_1 - 2$$
$$\underline{W}_3(8) = \underline{W}_3(7) = \begin{pmatrix}-2\\0\\-2\end{pmatrix} \longrightarrow d_3(\underline{x}) = -2x_1 - 2$$

$$d_{12}(\underline{x}) = d_1(\underline{x}) - d_2(\underline{x}) = -2(x_1 + x_2) + 2$$

Non - Linear Classifier *

Linear •

$h(\underline{x}) = \underline{v}^T \underline{x} + v_0$

$f(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2)$

$\frac{\partial f}{\partial \underline{v}} , \frac{\partial f}{\partial v_0}$

Non - Linear •

$y(\underline{x}) \equiv$ General Discriminent Function

$\underline{X}$

$y(\underline{X}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$

$f(\eta_1, \eta_2, s_1^2, s_2^2)$

$\eta_i = E\{ y(\underline{X}) \mid \omega_i \}$

$s_i^2 = E\{ y^2(\underline{X}) \mid \omega_i \}$

$s_i^2 =$ second order moment of y

$= \int_{-\infty}^{+\infty} y^2(\underline{x}) . p_i(\underline{x}) . d\underline{x}$

$\delta f \equiv$ variation of f due to variation of $y(\underline{x})$

$$\delta f = \frac{\partial f}{\partial s_1^2} \delta s_1^2 + \frac{\partial f}{\partial s_2^2} \delta s_2^2 + \frac{\partial f}{\partial \eta_1} \delta \eta_1 + \frac{\partial f}{\partial \eta_2} \delta \eta_2$$

$$\delta \eta_i = \int \delta y \cdot P_i(\underline{x}) \cdot d\underline{x}$$

$$\delta s_i^2 = \int 2\, y(\underline{x}) \cdot \delta y(\underline{x}) \cdot P_i(\underline{x}) \cdot d\underline{x}$$

مطلوبست :

$$\delta f = 0$$

$$y = a_1 \frac{sP_1(\underline{x})}{sP_1(\underline{x}) + (1-s)P_2(\underline{x})} + a_2 \frac{(1-s)P_2(\underline{x})}{sP_1(\underline{x}) + (1-s)P_2(\underline{x})}$$

$$s = \frac{\dfrac{\partial f}{\partial s_1^2}}{\dfrac{\partial f}{\partial s_1^2} + \dfrac{\partial f}{\partial s_2^2}}$$

$$a_i = -\frac{\partial f/\partial \eta_i}{2\, \partial f/\partial s_i^2}$$

- برای استفاده از σ^2 داریم :

$$s_x^2 = \sigma_x^2 + \eta_x$$

- مشابهت با قانون بیز

$$q_1(\underline{x}) \gtrless q_2(\underline{x})$$
$$q_1(\underline{x}) - q_2(\underline{x}) \gtrless 0$$
$$g(\underline{x}) = a_1 \cdot q_1(\underline{x}) + a_2 \cdot q_2(\underline{x})$$

- فرم تابع تمایز غیرخطی درجه دوم

$$\underbrace{\frac{1}{2}(\underline{x}-\underline{m}_1)^T \Sigma_1^{-1}(\underline{x}-\underline{m}_1) - \frac{1}{2}(\underline{x}-\underline{m}_2)^T \Sigma_2^{-1}(\underline{x}-\underline{m}_2) + \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|}}_{h(\underline{x})} \gtrless \ln\frac{P_1}{P_2}$$

$h(\underline{x}) \rightarrow$ تنها m_i , Σ_i مجهولاتش

• فرم کلی تابع تمایز کننده غیرخطی درجه دوم

$$h(\underline{x}) = \underline{x}^T [q_k] \underline{x} + \underline{v}^T \underline{x} + v_0 \gtrless_{\omega_2}^{\omega_1} 0$$

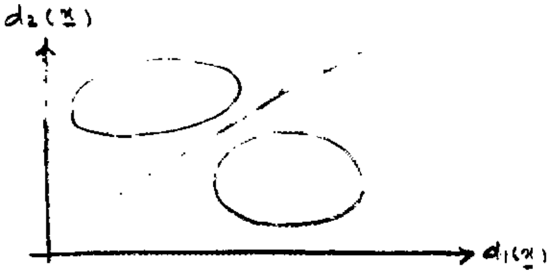
$$f(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2)$$

اگر فضا n بعدی باشد باید این تعداد پارامتر را تخمین زد:

$$\frac{n(n+1)}{2} + n + 1$$

• راه دیگر، نگاشت تابع درجه دوم، به توابع تمایز دهنده است و آنها را ازهم جدا کرد

$$d_1^2(\underline{x}) = (\underline{x} - \underline{m}_1) \Sigma_1^{-1} (\underline{x} - \underline{m}_1)$$
$$d_2(\underline{x}) = (\underline{x} - \underline{m}_2) \Sigma_2^{-1} (\underline{x} - \underline{m}_2)$$



تمرینات فصل 4

computer projects. 1, 2, 3, 4, 5(a, b, d, e, f), hard out
Problems : 1, 2, 3, 4, 5, 9
date : 1367 ,9 ,17

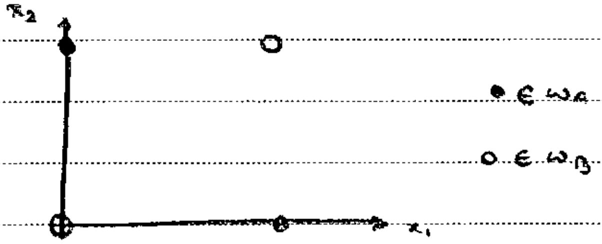
میان ترم:

date : 1367 ,9 ,17 _ , 9 ,24 !! ?
book : 2, 3, 4 → Fukunaga
classroom : ct. all.

Non - Linear Classifiers

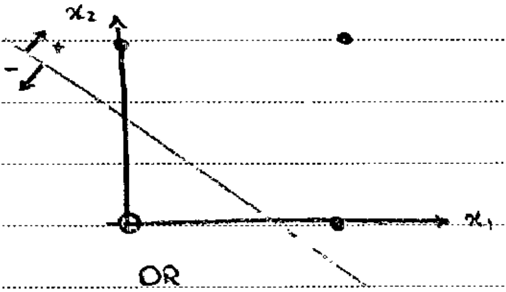
OR vs XOR ☑

x_1	x_2	class
0	0	False (−) : ω_B
0	1	True (+) : ω_A
1	0	True (+) : ω_A
1	1	False (−) : ω_B



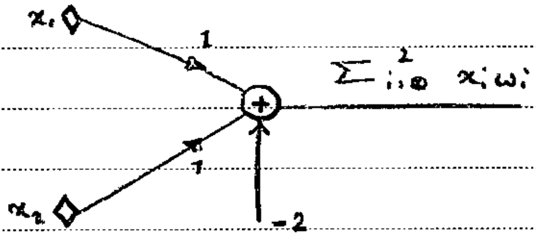
XOR

x_1	x_2	class
0	0	F : ω_B
0	1	T : ω_A
1	0	T : ω_A
1	1	T : ω_A



OR

☑ شبکه پرسپترون برای OR

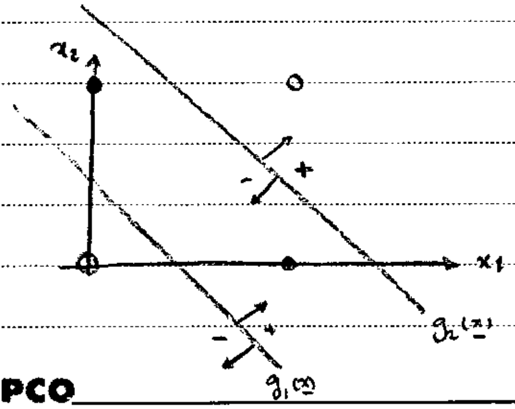


$$h(\underline{x}) = \underline{w}^T \underline{x} = [v_0 \quad \underline{v}]^T \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = v_0 + \underline{v}^T \underline{x}$$

$$h(\underline{x}) = x_1 + x_2 - \frac{1}{2} = 0$$

0	0	<0
0	1	>0
1	0	>0
1	1	>0

☑ شبکه پرسپترون برای XOR



$$f(x) = \sum_{i=0}^{2} w_i x_i$$

$$u(x) = \begin{cases} 1 & , x>0 \\ 0 & , x<0 \end{cases}$$

$$y_i = u_i(f_i(\underline{x})) \rightsquigarrow g_i(\underline{x})$$

step I :

y_1 موقعیت نسبت به پاره خط

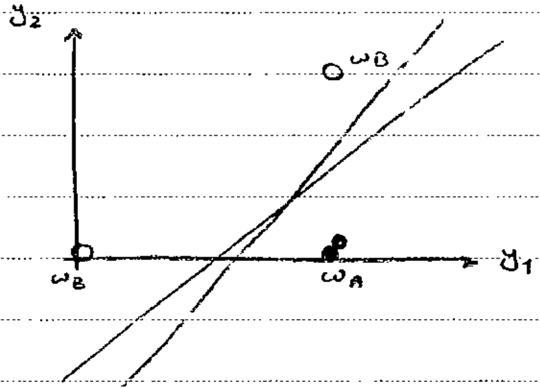
x_1	x_2	$y_1 = g_1(x)$	$y_2 = g_2(x)$
0	0	0	0
0	1	1	0
1	0	1	0
1	1	1	1

step II :

موقعیت نسبت به دو پاره خط

در کلاس : داخل خطوط، خارج خطوط

x_1	x_2	XOR
0	0	0
0	1	1
1	0	1
1	1	0



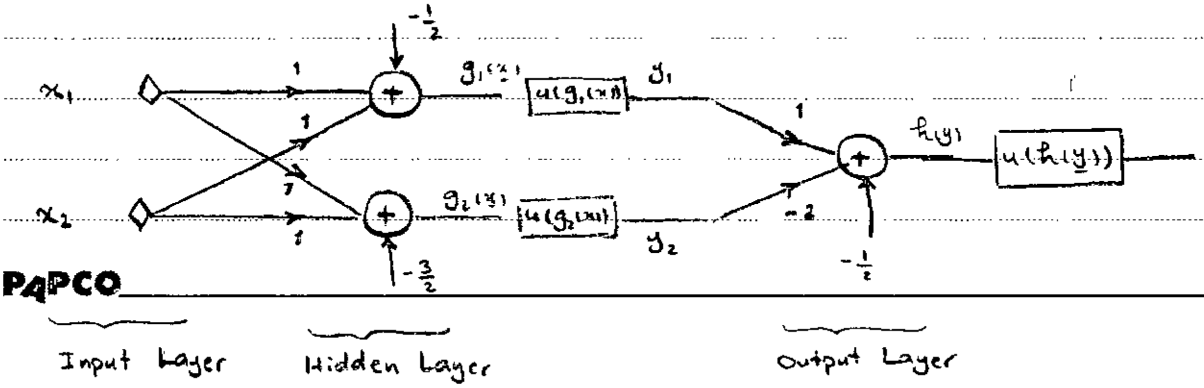
→ خطی جداپذیرند

عملاً از نگاشت استفاده شده است

y_i: انتخاب‌های → $u(f_i(\underline{x}))$
→ $\text{sgn}(f_i(\underline{x}))$
→ ...

⇒ firing function

پرسپترون دو لایه:



Two - Layer Perceptron

Two - Layer Feed Forward NN

$$g_1(\underline{x}) = x_1 + x_2 - \frac{1}{2}$$

$$g_2(\underline{x}) = x_1 + x_2 - \frac{3}{2}$$

$$h(y) = y_1 - 2y_2 - \frac{1}{2}$$

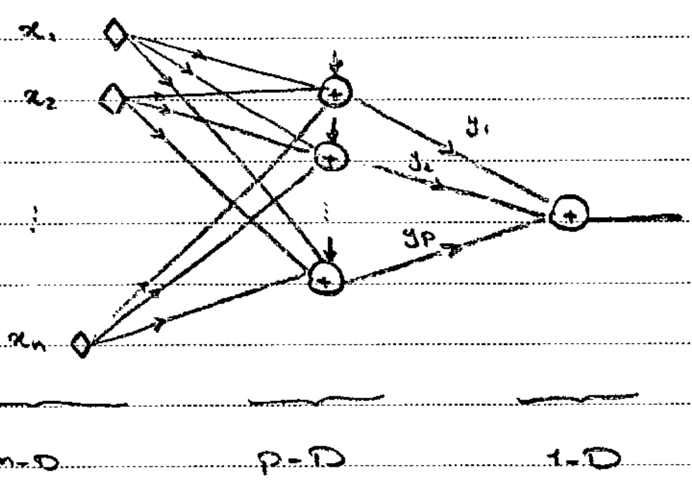
$$= u(g_1(\underline{x})) - 2\, u(g_2(\underline{x})) - \frac{1}{2}$$

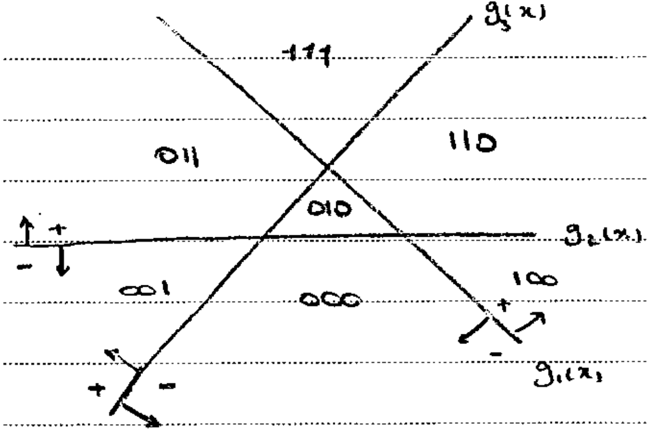
فضا را به یک مستطیل نگاشت می‌کند

- Multi - Layer Perceptron (MLP)

n - D , 2 - class

hyper - cube !





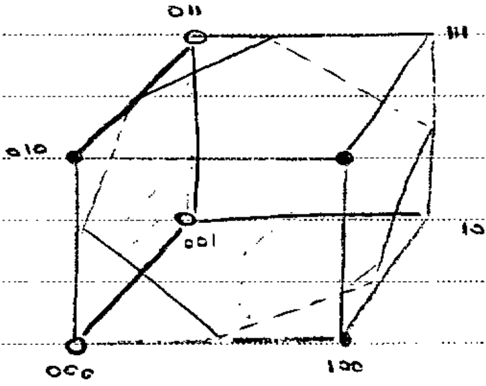
☑ فضای سه بعدی مسئله XOR

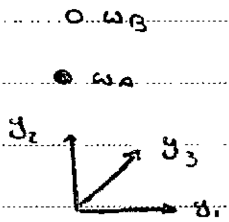
x_1 \ x_2x_3	00	01	11	10
0		1		1
1	1		1	

$g_1(x) \longrightarrow$ بیت اول

$g_2(x) \longrightarrow$ بیت دوم

$g_3(x) \longrightarrow$ بیت سوم





Generalization

$\underline{x}$

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} f(g_1(x)) \\ f(g_2(x)) \end{bmatrix}$$

$f(..) \equiv$ firing function $\equiv$ activation function

generalized linear function:

$f_i : R^n \longrightarrow R^k$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}$$

$$g(\underline{x}) = w_0 + \sum_{i=1}^{k} w_i f_i(\underline{x})$$

2nd Order

$$g(\underline{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n-1} \sum_{m=i+1}^{n} w_{im} x_i x_m + \sum_{i=1}^{n} w_{ii} x_i^2$$

n-D

$$\underline{y} = \underline{g}(x) = [x_1 \quad x_2 \quad x_1 x_2 \quad x_1^2 \quad x_2^2]$$

2-D

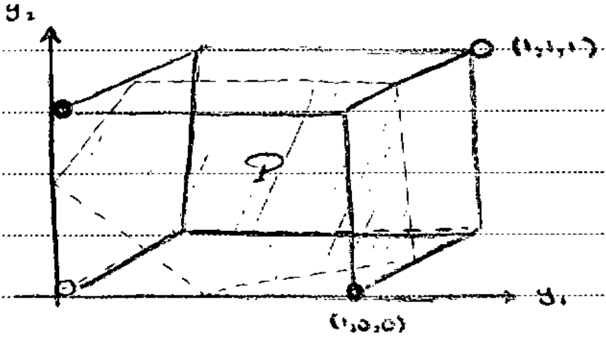
r- order
n- D $\Rightarrow$ $x_1 \; x_2 \; \dots \; x_n^{r_n}$

$$k = \frac{(n+r)!}{n! \times r!}$$

XOR ☑

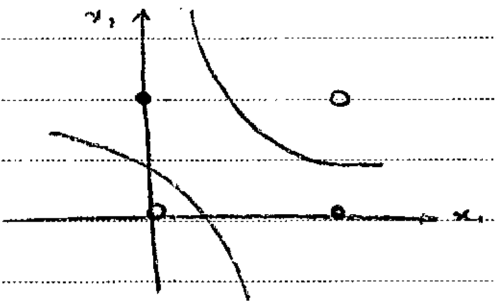
$$\underline{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{pmatrix}$$

$\underline{x} = (x_1, x_2)$	(0,0)	(0,1)	(1,0)	(1,1)
$\underline{y} = (y_1, y_2, y_3)$	(0,0,0)	(0,1,0)	(1,0,0)	(1,1,1)
	ω_B	ω_A	ω_B	ω_A



$P: \; h(y) =$
$y_1 + y_2 - 2y_3 - \frac{1}{4} = 0$

$(0,0,0) \rightarrow h(0,0,0) < 0$ نسبت صفر



$h(\underline{y}) = y_1 + y_2 - 2y_3 - \frac{1}{4} = 0$
$g(\underline{x}) = x_1 + x_2 - 2x_1 x_2 - \frac{1}{4} = 0$

(RBF) Radial Basis Function Networks •

$\Phi(.) = f(\| \underline{x} - \underline{c}_i \|)$ $\hookrightarrow$ distance function

$$g(\underline{x}) = w_0 + \sum_{i=1}^{k} w_i \cdot \Phi_i(.)$$

برای $f(y)$ انتخاب‌های متفاوتی وجود دارد. از آنجایی که با فاصله سروکار دارد به آن RBF می‌گویند.

$$\star \quad f(\underline{x}) = \exp\left(-\frac{1}{2\sigma_i^2}\|\underline{x}-\underline{c}_i\|^2\right)$$

$$g(\underline{x}) = w_0 + \sum_{i=1}^{k} w_i \exp\left(-\frac{(\underline{x}-\underline{c}_i)^T(\underline{x}-\underline{c}_i)}{2\sigma_i^2}\right)$$

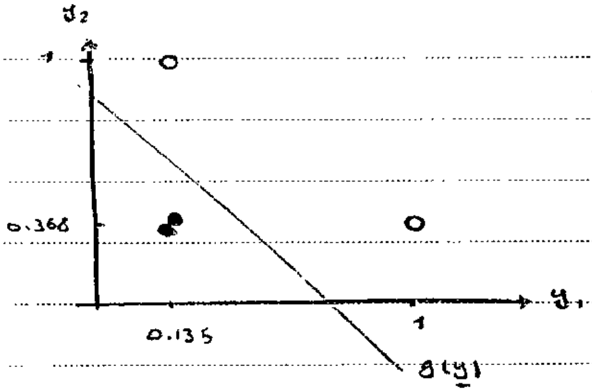
$$\star \quad f(\underline{x}) = \frac{\sigma^2}{\sigma^2+\|\underline{x}-\underline{c}_i\|^2}$$

XOR ☑

$$\underline{y} = y(\underline{x}) = \begin{bmatrix} \exp(-\|\underline{x}-\underline{c}_1\|^2) \\ \exp(-\|\underline{x}-\underline{c}_2\|^2) \end{bmatrix}$$

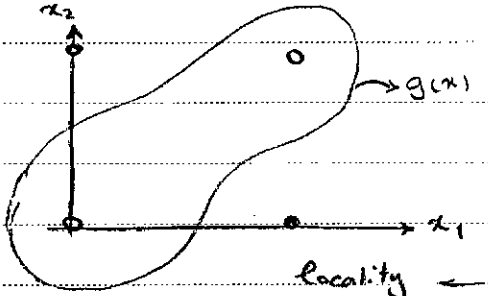
$$c_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad c_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

x_1	x_2	y_1	y_2
0	0	0.135	1
0	1	0.368	0.368
1	0	0.368	0.368
1	1	1	0.135



$$g(\underline{y}) = y_1 + y_2 - 1 = 0$$

$$g(\underline{x}) = \exp\left(-(\underline{x}-\binom{1}{1})^T(\underline{x}-\binom{1}{1})\right) + \exp\left(-(\underline{x}-\binom{0}{0})^T(\underline{x}-\binom{0}{0})\right) - 1 = 0$$



این تابع به seed وابسته هستند (c_i) ← locality

می‌توان مقادیر c (مراکز)، σ (پهنه) و w را با یادگیری بدست آورد.

* Clustering ⟶ تعیین : c_i ها

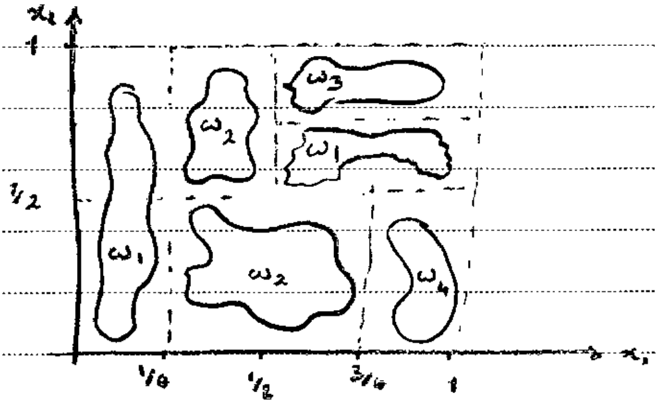
* Gradient ⟶ $J(\underline{x}, \underline{w}, \underline{c}, \sigma)$

$$\underline{W}(k+1) = \underline{w}(k) - c\,\nabla_{\underline{w}} J$$

$$\underline{c}(k+1) = \underline{c}(k) - c\,\nabla_{\underline{c}} J$$

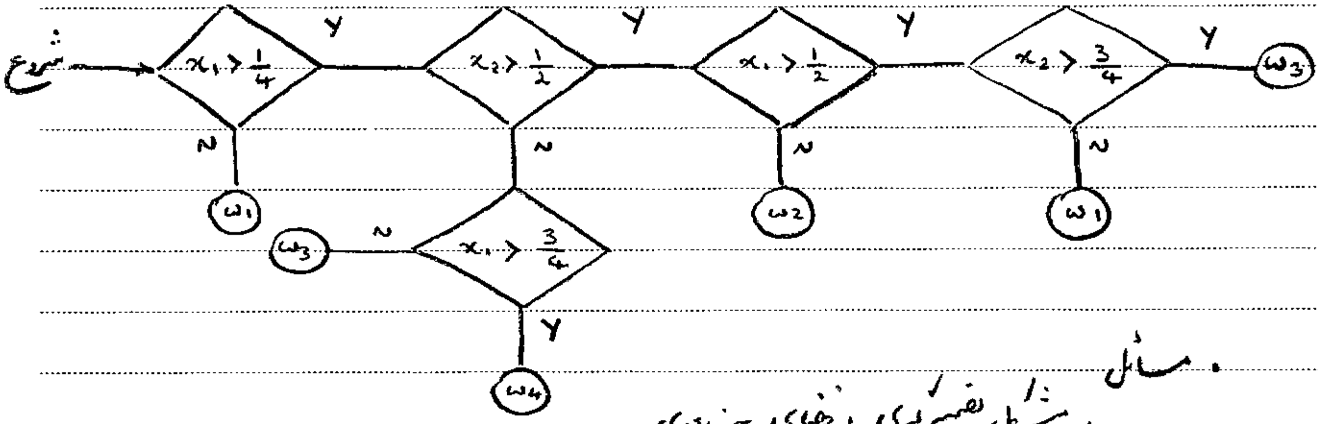
$$\sigma(k+1) = \underline{c}(k) - c\,\nabla_{\sigma} J$$

* درختهای تصمیم (Decision Trees) ✓



4-class

2-D



مسائل:

* مشکل تصمیم گیری درختهای چندبعدی
* کدام ویژگی برای تصمیم گیری استفاده شود.
* نامگذاری برگها
* هرس کردن درختهای بزرگ
* Missing Feature

فصل 6) Non-Parametric Density Estimation *

$P_i(\underline{x}) \qquad i = 1, \dots, M$

* گاهی اوقات فرض می‌کنیم که توزیع ما چیست و پارامترهای آنها تخمین می‌زنیم

مثلاً μ , Σ برای توزیع نرمال

* گاهی اوقات تنها از روی N نمونه مدل ساخته می‌شود و می‌خواهیم از روی

نمونه‌ها توزیع را بیاورد.

1-D.

$p(\underline{x}) \rightsquigarrow$ معلوم

$R \equiv$ Region

$$\text{Prob}\{x \in R\} \equiv P = \int_R p(x) \cdot dx$$

n-D

$$\text{Prob}\{\underline{x} \in R\} \equiv P = \int_R p(\underline{x}) \cdot d\underline{x}$$

n-integral

$p(\underline{x})$ is continuous over R

$$P = \int_R p(\underline{x}') \cdot d\underline{x}' \approx p(\underline{x}) \cdot V$$

$V \equiv$ volume enclosed by R

$\underline{x} \equiv$ a point in R

N samples

$\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$

probability that "K" of "N" samples fall in R is estimated by:

$$P \approx \frac{K}{N}$$

- از ترکیب دو رابطه داریم:

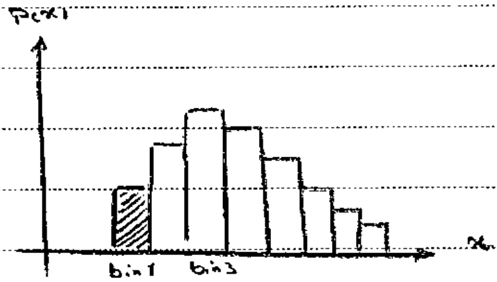
$$\hat{p}(\underline{x}) . V = \frac{K}{N} \Rightarrow \hat{p}(\underline{x}) = \frac{K}{N . V} = \frac{K / N}{V} = \frac{\hat{P}}{V}$$

$$\Rightarrow \hat{p}(\underline{x}) = \frac{\int_R p(\underline{x}) . d\underline{x}}{\int_R d\underline{x}}$$ ~ smooth version of $p(\underline{x})$

- با کوچک کردن فاصله

* $\hat{p}(\underline{x}) \longrightarrow p(\underline{x})$
* ممکن است در بازهٔ در نظر گرفته هیچ نمونه‌ای نباشد ← حالت spike
* دو محدودیت فوق یک dilemma می‌سازد.
* به این فاصله کوچک یک bin یا cell می‌گویند.

* روش Histogram



- $\underline{x}_1, \dots, \underline{x}_N$

bins

Area under the j^{th} bin (interval) =

Fraction of total # of samples (N) fall in that region =

$$\frac{K_j}{N}$$

Height of the density $= \frac{\text{area}}{\text{width}} = \frac{K_j / N}{W_j} = \frac{K_j}{N . W_j}$

- تعداد Bin
  - * زیاد ← تخمین زمخت از تابع توزیع
  - * کم ← حالت spike و درگیر شدن با جزئیات
  - * Rule of thumb

N samples

$\# \text{ of interval} = \sqrt{N}$

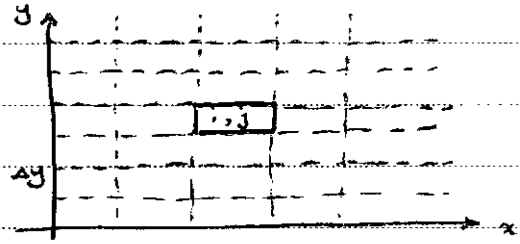
* برای مینیمم کردن خطای مربعی

$E\{(p(x) - \hat{p}(x))^2\}$

$\# \text{ of interval} = \sqrt[3]{N}$

- 2-D

$P(\underline{x})$ is continuous ove R



- 5-D

$n = 5$

10 intervals

10^5 histogram bins → lots more samples requires

- می‌توان از بازه‌های نابرابر استفاده کرد.

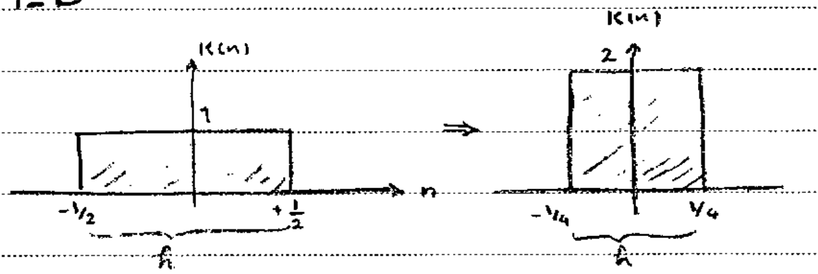
Parzen Window *

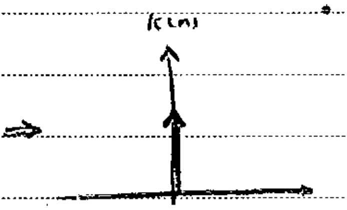
- در این روش، هر نمونه را یک Kernel احاطه می‌کند که دور آن را می‌گیرد.
برای اینکه در هر نقطه اندازه توزیع تخمین زده شود، مقادیر Kernel های موجود در
آن نقطه را با هم جمع می‌کنند.

Convolution

$$x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau)\, h(t-\tau)\, d\tau$$

1-D





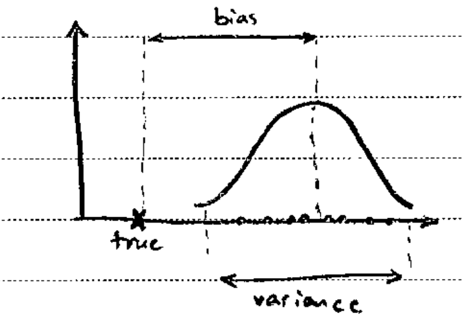
$h \to 0 \Rightarrow K(n) \to \delta(n)$

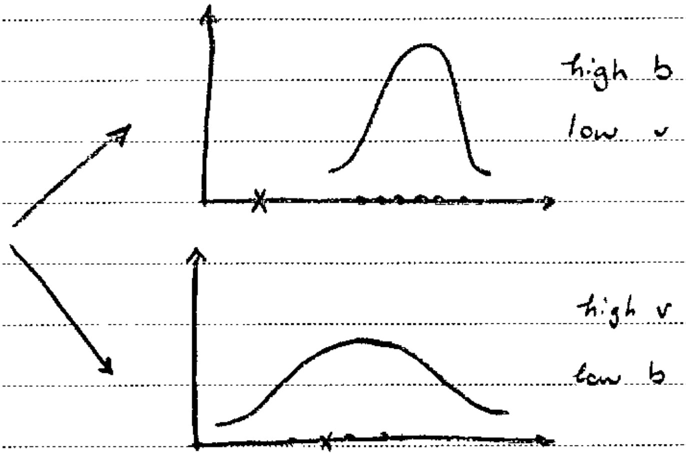
$x(t) * \delta(t) = x(t)$

← یعنی تخمین به سمت مقدار واقعی می‌رود.

- Bias: how close is the estimate to the true value?

Variance: how much does the estimate change for different runs?

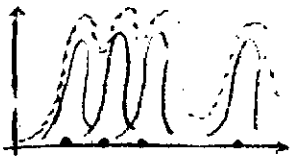




Trade off: bias - variance

- در روش Parzen ، به صورت زیر یعنی kernel ابتدا باید
برای یک sample ابتدا یک Kernel قرار می‌دهیم
نتیجه kernel های هر نقطه ، با هم جمع کردن تخمین بدست می‌آید.
این تابع می‌تواند مستطیلی ، گوسی و ... باشد.

- kernel → باریک : تخمین Spikey
→ پهن : تخمین Smooth

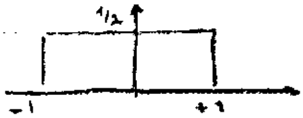


$$MSE\,(P_{KDE}) = Bias^2 + Variance$$

- انواع kernel های مختلف برای یک بعد

I) Rectangular

$$k(x) = \begin{cases} \frac{1}{2} & |x| < 1 \\ 0 & ,|x| > 1 \end{cases}$$



II) Triangular

$$K(x) = \begin{cases} 1 - |x| & ,|x| < 1 \\ 0 & ,o.w \end{cases}$$

III) Biweight

$$K(x) = \begin{cases} \frac{15}{16}\,(1 - x^2)^2 & ,|x| < 1 \\ 0 & ,o.w \end{cases}$$

IV) Normal

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Subject:
Year. Month. Date. 56

V) Bartlett - Epanechnikov

$$K(x) = \begin{cases} \frac{3}{4}\left(1-\frac{x^2}{5}\right)/\sqrt{5} & , |x| < \sqrt{5} \\ 0 & , \text{o.w} \end{cases}$$

• فضای چندبعدی

I) Normal

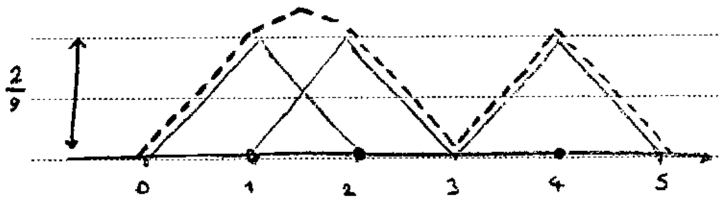
$$k(\underline{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\underline{x}^T\underline{x}\right\}$$

II)

$$k(\underline{x}) = (1 - \underline{x}^T\underline{x})\frac{(n+2)}{2C_p} \quad \text{for } |x| < 1$$

$$C_p = \pi^{n/2} / \Gamma\left(\left(\frac{n}{2}\right)+1\right)$$ ~ volume of sphere in n-D

☑



base = 3 مثلث

$\hat{p}(\underline{x})$

area of each triangle = $\frac{1}{3}$

height = $\frac{2}{9}$

• برای انتخاب پهنای باند

Minimize Integral of Square Error — * استفاده از MISE

توزیع نرمال و kernel نرمال

$h_{opt} = 1.06 \times \sigma \times N^{-1/5}$ — * با استفاده از h ثابت

$$h_{opt} = \underset{h}{\arg\min}\left\{\text{MISE}(P_{KDE}(x))\right\} = \underset{h}{\arg\min}\left\{E| \qquad \right\}$$

PAPCO

Subject:
Year. Month. Date. ( )

* یک معیار دیگر A است.

$$A = \min \left\{ \sigma , \frac{IQR}{1.34} \right\}$$

$$h_{opt} = 0.9 \times A \times N^{-1/5}$$

IQR = بازه‌ای از پراکندگی داده‌ها که بین 0.25 تا 0.75 داده‌ها درآن باشد

* h های مختلف در راستاهای گوناگون ممکن است مناسب باشد.

* kernel را به صورت ضرب تعدادی kernel مختلف می‌توان نوشت که هرکدام ممکن است h های گوناگون داشته باشند. هرکدام از این زیر kernel ها درجه یک هستند. فرض استقلال متغیرها در این روش مهم نیست.

* می‌توان از معیارهای دیگری نیز استفاده کرد. به جای MISE، سعی در کمینه KL کند.

- اگر فرض استقلال مولفه‌های P(x) را داشته باشیم می‌توان آنها را اینگونه نوشت:

$$P(\underline{x}) = \prod_{i=1}^{n} P(x_i)$$

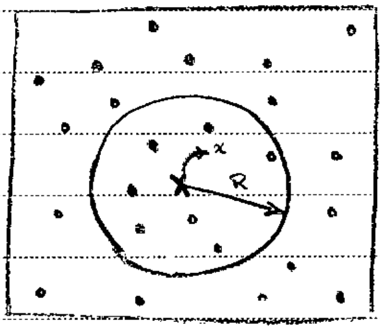
☑ Naive Bayes Classifier

* K Nearest Neighbor

- پراکندگی نقاط در همه جا یکسان نیست.
- سعی می‌کنیم اندازه bin را با توجه به حجم نمونه‌ها در نظر بگیریم.
- در Parzen Window، اندازه bin ثابت است و می‌بینیم چند نمونه در آن می‌افتد.
- در KNN، تعداد نمونه‌ها در هر bin ثابت است و می‌بینیم اندازه آن چقدر باید باشد.

$P(x) = \frac{k}{N.V}$

$n = D$



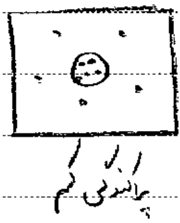
$P(x) = \frac{k}{N \cdot C_n \cdot R_k^n(x)}$

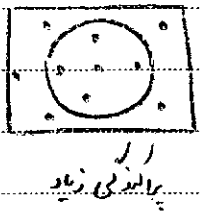
$C_n = \frac{\pi^{n/2}}{(n/2)!} = \frac{\pi^{n/2}}{\Gamma(n/2+1)!}$

$C_1 = 2$

$C_2 = \pi$

$C_3 = \frac{4\pi}{3}$





* حساس به Local Noise

* حالت Spikey ← چون تابع R(x) مشتق پذیر نیست.

* هرچقدر نمونه‌ها بیشتر باشد، تخمین بهتر است

* K با توجه به تعداد نمونه‌ها مقدار opt دارد ← K< opt : زودحش

K> opt : بی‌خیال

* برای چنین توزیع مانند پیز

$l(x) = \frac{P_1(x)}{P_2(x)}$

$P(x|\omega_i) = \frac{K_i}{N_i.V}$

$P(\omega_i) = \frac{N_i}{N}$

$P(\omega_i|x) = \frac{P(x|\omega_i)\,P(\omega_i)}{P(x)} = \frac{K}{N}$ ⤳ KNN Classifier

Subject:
Year. Month. Date. ( )

● KNN Classifier

- اول k تا از نزدیک ترین همسایه های x را پیدا می کنیم
- می بینیم هر کدام از این همسایه ها متعلق به کدام کلاس است.
- بیشترین تعداد را به عنوان کلاس x مشخص می کنیم

● KNN ← Lazy Learner

- نیاز به تمام داده های یادگیری دارد تا ورودی جدید را طبقه بندی کند.
- هر موقع لازم بود طبقه بندی را انجام می دهد.
- برخلاف Statistical PR یا Neural PR
- نیاز به حافظه زیاد
- حجم محاسباتی کم
- Curse of Dimentionality
- k بزرگ ← جزئیات از بین می رود
  محاسبات بیشتر
- k کوچک ← Smoothness از دست می رود.

P4PCO

Independent Bayes , Naïve Bayes , Idiot's Bayes *

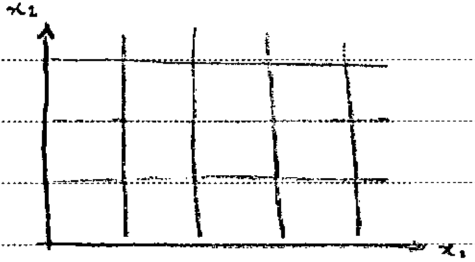
$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \longrightarrow X_i \quad \text{باهم مستقل}$$

$$P(\underline{x}) = \prod_{i=1}^{n} P(x_i)$$

بیشترو .

$x_i \longrightarrow L$ bin

$x \longrightarrow n \times L$ bin $\longrightarrow$ مستقل



$\longrightarrow L^n \longrightarrow$ غیرمستقل

independent model .

$$P(\underline{x}) \sim \left\{ \prod_{r=1}^{n} \frac{n(x_r) + 1/C_r}{N(r) + 1} \right\}^{B}$$

$x_r \equiv r^{th}$ Component of $\underline{x}$

$n(x_r) \equiv$ # of samples with value x_r on variable r

$N(x_r) \equiv$ # of observation on variable r (due missing value)

$C_r \equiv$ # of cells on variable r

$B \equiv$ "Association factor" $\hookrightarrow$ non-redundant information با تناسب

$B = 1 \longrightarrow$ روش معمول

$B < 1 \longrightarrow$ معمولاً $\longrightarrow$ 0.5 , 0.8 , ...

Lancaster Model

$P(\underline{x}) \longrightarrow$ می‌خواهیم تخمین بزنیم

اجازه می‌دهیم که تا یک حدای بین مثلاً ۲ interaction برقرار باشد

n-D

$\underline{X}$

$$P(\underline{x}) = \left\{ \sum_{\substack{i,j\\ i<j}} \frac{P(x_i, x_j)}{P(x_i).P(x_j)} - \left[\binom{n}{2} - 1 \right] \right\} P_{ind}(\underline{x})$$

$$P_{indep}(\underline{x}) = \prod_{k=1}^{n} P(x_k)$$

$S \equiv$ order of interaction $= 2$

$$P(x_i, x_j) = \frac{n(x_i, x_j) + \frac{1}{c_i \times c_j}}{N(i,j) + 1}$$

$n(x_i, x_j) \equiv$ # of samples with value x_i, x_j

$N(x_i, x_j) \equiv$ # of observation on variable i & j

$c_r \equiv$ # of cells in variable r

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$	$\begin{bmatrix} x_1 \\ * \end{bmatrix}$	$\begin{bmatrix} * \\ x_1 \end{bmatrix}$		☑
✓	✓	✓	n	
✓	x	x	N	

Maximum Weight Dependence Tree *

2-D

$$P_i(\underline{x}) = P(\underline{x} \mid \omega_i) = \frac{P(\underline{x}; \omega_i)}{P(\omega_i)}$$

$$\Rightarrow P(\underline{x}; \omega_i) = P(\omega_i) \times P(\underline{x} \mid \omega_i)$$

$t \equiv$ tree

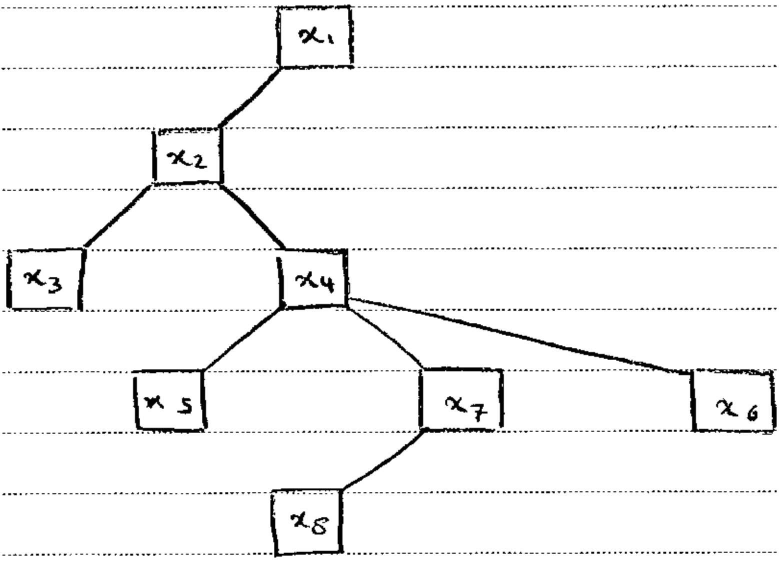
$$p^t(\underline{x}) = \prod_{i=1}^{n} P(x_i \mid x_{j(i)})$$ پدری ندارد که ریشه x_1 باشد

$$= \prod_{i=1}^{n} P(x_{m_i} \mid x_{j(m_i)})$$

$j(i) = \text{parent}(i)$

☑

$$p^t(\underline{x}) = P(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2) \cdot p(x_4 \mid x_2) \cdot p(x_5 \mid x_4) \cdot p(x_6 \mid x_4) \cdot p(x_7 \mid x_4) \cdot p(x_8 \mid x_7)$$

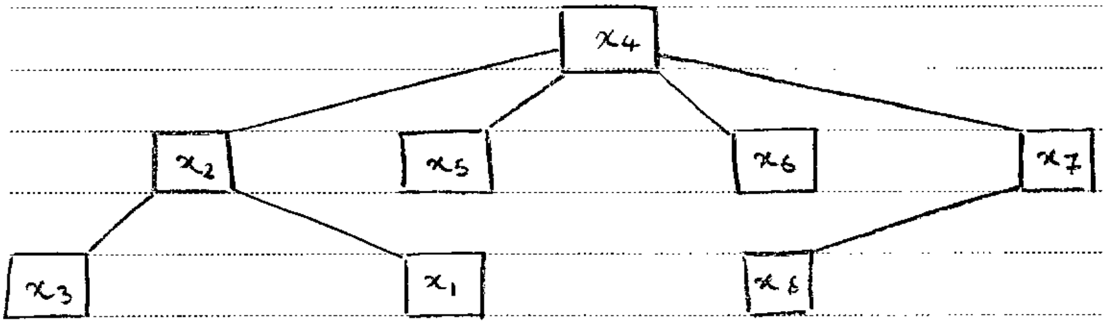


- به دنبال درختی هستیم که تخمین خوبی $p^t(\underline{x})$ باشد.

$x_1 \longrightarrow$ root : $p(x_1 \mid x_0) = x_1$

ادامه مثال ☑

$$P^t(\underline{x}) = P(x_4) . P(x_1 | x_2) . P(x_3 | x_2) . P(x_2 | x_4)$$
$$. P(x_5 | x_4) . P(x_6 | x_4) . P(x_7 | x_4) . P(x_8 | x_7)$$



- حداکثر تعداد درختها

n^{n-2}

- اگر هر متغیر L تا سطح داشته باشد.

$\# \text{ of cells} = L(L-1)(n-1) + L$

- معیاری مورد نیاز است تا نزدیکی تخمین را اندازه‌گیری کند.

معمولاً از شاخص KL استفاده می‌کند ← Kullback Leibler

$$D_{KL}(p(\underline{x}), P^t(\underline{x})) = \int p(\underline{x}) . \log\left(\frac{P(x)}{P^t(x)}\right) . d\underline{x}$$

$$D_{KL} \longrightarrow > 0$$
$$\searrow = 0 \Rightarrow p(\underline{x}) = P^t(\underline{x}) \Longrightarrow$$ بهترین تخمین

$$D_{KL}(P(\underline{x}), P^T(\underline{x})) \leqslant D_{KL}(P(\underline{x}), P^t(\underline{x})) \quad \forall t \neq T$$

این تابع cross entropy را کاهش می‌دهد. مقدار entropy عدم قطعیت موجود در تابع را نشان می‌دهد.

Entropy ●

$\{X_1, X_2, \ldots, X_m\}$

$P_i = \text{Prob}\{X_i\} \quad i = 1, \ldots, m$

Entropy: $H_2 = -\sum_{i=1}^{m} P_i \log_2 P_i$ ~ bits

☑

$X_0, X_1, \ldots, X_7$

$P_i = \frac{1}{8}$

$H = 3 \longrightarrow$ # of bits to represent symbols

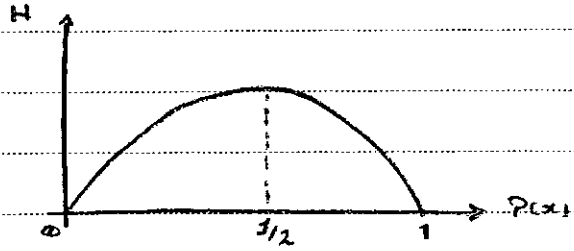
☑

$X_0, X_1, \ldots, X_7$

$P_i = 0 \quad i = 0, \ldots, 6$

$P_7 = 1$

$H = 0 \longrightarrow X_7$ certainely occures = no uncertainaty



●

$P(x)$, $q(x)$

x ~ discrete

$$D_{KL}(P(x), q(x)) = \sum_{x} q(x) \, Ln \frac{q(x)}{P(x)}$$

- خصوصیات D_{KL}

$D_{KL} \geq 0$

$D_{KL} = 0 \iff q(x) = p(x)$

- Mutual Information (MI)

different variables

$p(x)$

$q(y)$

$I(p,q) = H(p) - H(p|q)$

مقدار عدم قطعیتی است که از مقدار اول کاسته می‌شود وقتی که ما یک متغیر را داریم.

$$I(p,q) = H(p) - H(p|q)$$
$$= \sum r(x,y) \log_2 \frac{r(x,y)}{p(x) \cdot q(y)}$$

$r(x,y) \equiv$ joint distribution of x, y

MI نشان می‌دهد که متغیر چقدر از هم مستقل هستند.

$MI \geq 0$

$MI = 0 \iff$ independance

-

$p^t(x) = \prod_{i=1}^{n} P(x_i \mid x_{j(i)})$

هر شاخه درخت دارای وزن است که این وزن معادل MI بین پدر و فرزند است.
درختی مناسب است که مجموع وزن‌های آن بیشینه باشد.

$W = \sum_{i=1}^{n} I(X_i , X_{j(i)})$

$W \sim \max \Longrightarrow D_{KL} \sim \min$

- الگوریتم تولید درخت

max weight dependence tree (MWDT)
max weight spanning tree (MWST)

یک گراف کامل تشکیل می‌دهد که برچسب آن MI بین دو متغیر احتمالی باشد،
بین آنها MST پیدا کرد.

$$\left.\begin{array}{l} \text{find all weights} = \binom{n}{2} = \frac{n(n-1)}{2} \\ \text{find minimum spanning tree} : O(n^2) \end{array}\right| \Rightarrow O(n^2)$$

✏ تمرین سری چهارم

Density Estimation
1. Histogram
2. Parzen Method
    - * Rectangular
    - * Gaussian
    - * Triangular
    - * Product Form
3. KNN Method
4. Independance Tree

Subject:
Year. Month. Date. ( )

I. Generate Data
- * Normally Distributed 1D: $\{\mu \quad \sigma^2\}$ 2D: $\{\underline{m} \; [\Sigma]\}$
- * Uniform

II. Do Estimation

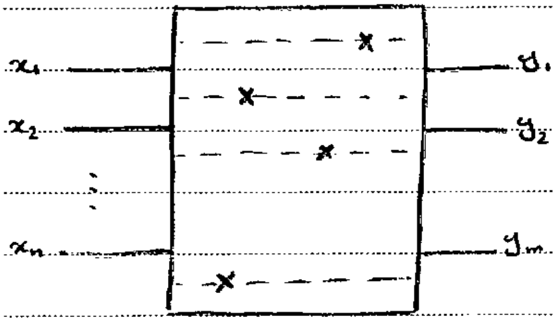
III. Visualize Data & PDF {original , estimated}

IV. Provide Interface

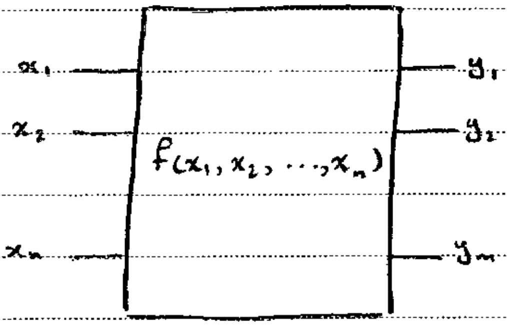
PAPCO

Feature Reduction *

n-D $n \times 1$ $\underline{X}$ $\underline{x}$
m-D $m \times 1$ $\underline{Y}$ $\underline{y}$) $m < n$



Feature Selection



Feature Extraction

$$\underset{m\times1}{\underline{y}} = [a]^T \underset{n\times1}{\underline{X}}$$

$$[a] = n \times m$$

• شاخص ارزیابی برتری یک روش نسبت به دیگری

feature selection ⇒ subset
feature extraction ⇒ function

$S_m \equiv$ all subsets

$$J(\tilde{S}_m) = \max_{s \in S_m} J(s)$$

$A \equiv$ all transformation

$$\tilde{J}(\tilde{A}) = \max_{A \in A} J(A(\underline{x}))$$

• تعداد روشهای ممکن

n feature

m select

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!}$$

$n = 25$ $m = 10$ ☑

$\# = 3,268,760$

Feature Selection Criterion •

1. Design Classifier → ...

2. $\underline{n\text{-}D}$ N_1 sample N_2

$\underline{m\text{-}D}$ $\underline{N_1}$ $\underline{N_2}$

$P_1(\underline{y})$ $P_2(\underline{y})$

Use Distance Functions ...

* Divergance

$$J_D(\omega_1, \omega_2) = \int_{-\infty}^{+\infty} \left[P(\underline{x}|\omega_1) - P(\underline{x}|\omega_2) \right] \times \ln \frac{P(\underline{x}|\omega_1)}{P(\underline{x}|\omega_2)} \times d\underline{x}$$

☑

class 1: $N(\underline{m}_1, [\Sigma]_1)$

class 2: $N(\underline{m}_2, [\Sigma]_2)$

$$J_D(\omega_1,\omega_2) = \frac{1}{2}(\underline{m}_2 - \underline{m}_1)^T \left([\Sigma]_1^{-1} + [\Sigma]_2^{-1} \right) (\underline{m}_2 - \underline{m}_1)$$
$$+ \underset{\text{trace}}{Tr} \left\{ [\Sigma]_1^{-1} [\Sigma]_2 + [\Sigma]_1 [\Sigma]_2^{-1} - 2[I] \right\}$$

* Chernoff

$$J_C(\cdot) = -\log \int p^{s}(\underline{x}|\omega_1) \cdot p^{(1-s)}(\underline{x}|\omega_2) \cdot d\underline{x}$$

* Bhattacharyya

$$J_B(\cdot) = -\log \int [p(\underline{x}|\omega_1) \cdot p(\underline{x}|\omega_2)]^{1/2} \cdot d\underline{x}$$

* Patrick - Fisher

$$J_P(\cdot) = \int [p(\underline{x}|\omega_1) P_1 - p(\underline{x}|\omega_2) P_2]^{2} \cdot d\underline{x}$$

* KL

• روشهای جستجوی زیرمجموعه‌ی Branch & Bound

* روش باید خاصیت monotonic داشته باشد، یعنی:

$X \subset Y \Longrightarrow J(X) < J(Y)$

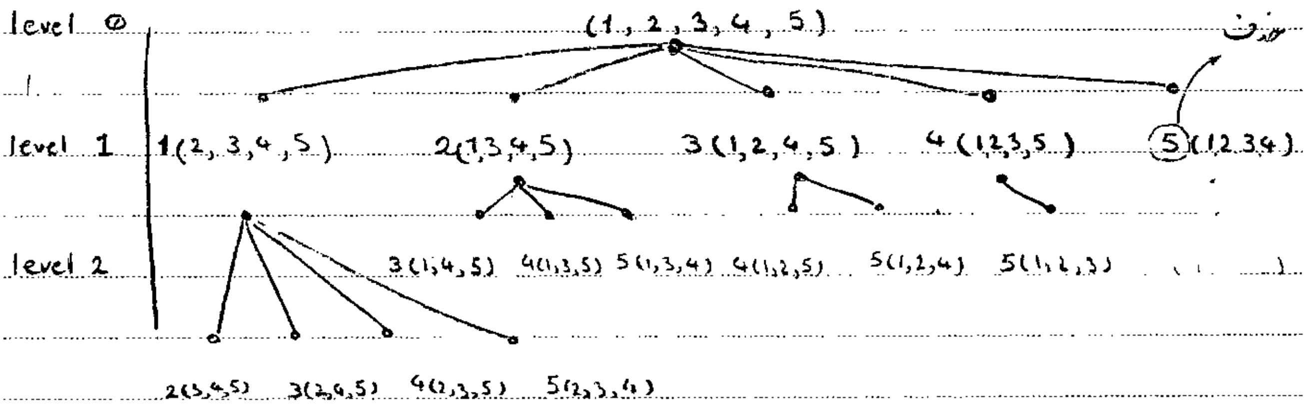
یعنی هرچه feature اضافه کنیم، مقدار J زیاد شود.

* روش optimum است.
* درخت تشکیل می‌دهد.
* از نوع Top - Down است، یعنی از کل feature ها شروع می‌کند سپس با بقیه ادامه دهد.
* شاخه‌های تکراری subset ها را در خود دارد تولید نمی‌کند.

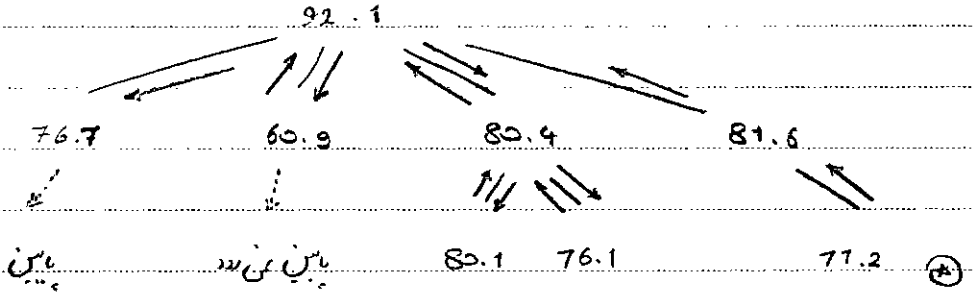
5 features ☑

select 3 features

construct a tree



criteria function اعمال مقدار



start @ ⊛ : $(1, 2, 3) \longrightarrow J = 77.2$

back track : $(1, 2, 5) \longrightarrow J = 80.1$

$$J = (\underline{m}_1 - \underline{m}_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\underline{m}_1 - \underline{m}_2)^T$$

علت اینکه پایین نمی‌رود (و ادامهٔ زیردرختها را نمی‌سازد) اینست که نود والد آنها مقدار J کمتر از مقدار بهینه ما باشد ، پس طبق خاصیت monotonic فرزندان آنها حتماً کمتر از خودشان می‌باشند و نمی‌توانند J را بهتر کنند

Feature Subset Selection → Search Strategy
(FSS) ↘ Objective Function (Criterion)

• انواع تابع ارزیابی

1. Filter → Information Content
2. Wrapper → Pattern Classifier

• انواع فیلترها

* فاصله کلاسها را مشخص کند. ← اقلیدسی
↙ ماهالانوبیس
↓ $S_w^{-1} \cdot S_B$

* به همبستگی ویژگیها بستگی دارد. ← خطی
↙ غیرخطی

• استراتژی های جستجو

..., Branch & Bound, Exhaustive ← Exponential *
Sequential *
..., Simulated Annealing, Genetic ← Randomized *

• Naïve Sequential Feature Sel.

$X_1, \ldots, X_n$
select m feature
use criteria function J
rank features regarding to J and indepedently : $J(X_i)$
select m upper feature
suitable for all independant feature set

* اگر از هم مستقل نباشند این جواب مناسب نیست ولی اگر ۲ تا ۲ feature ها از هم مستقل باشند، بهترین جواب را بدست می دهد.

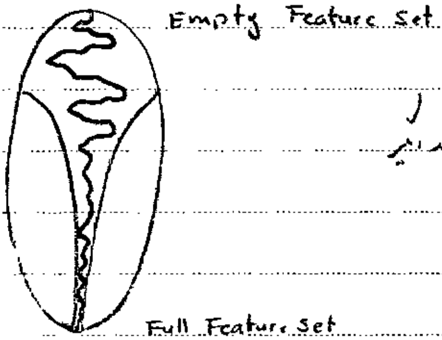
• Sequential Forward Sel (SFS)

* انتخاب یک زیرمجموعه از مجموعه feature ها به صورت Bottom-Up

* الگوریتم

1. Start with empty set Y_0
2. Add next best feature maximizing $J(Y_0+X_i)$

* شکل



* اشکال آن این است که وقتی یک feature انتخاب شد دیگر نمی توان آن را حذف کرد.

* در هر جای الگوریتم وقتی به تعداد مورد نظر feature داشتیم می توان الگوریتم را قطع کرد.
* برای مسائلی که تعداد کمی feature مورد نیاز است مناسب است

• Sequential Backward Sel (SBS)

* انتخاب یک زیرمجموعه به صورت Top-Down

* الگوریتم

1. Start with full set $Y_0 = X$
2. Remove the feature causing less decrease in criteria

* اشکال: قابلیت بازگشت feature حذف شده وجود ندارد
* برای مسائل با تعداد feature مورد نیاز زیاد

(LRS) Plus -L Minus R Sel. •

* در هر مرحله L تا اضافه و R تا را حذف می‌کنیم.

1. Depend on L or R start Y = Φ / X
2. Add L and remore R ,...

(BDS) Bidirectional Search •

SFS + SBS *

* اگر می‌خواهد Feature اضافه کند باید در نظر بگیرد که اگر آنها حذف نکرده باشد.

(SFFS) Sequential Floating Sel. •

(chapter 9)

Feature Extraction for Signal Representation *

- هدف پیدا کردن بهترین feature ها برای نمایش سیگنال و کمترین از دست دادن اطلاعات

$$\underline{X} = Y_1 \underline{u}_1 + Y_2 \underline{u}_2 + \cdots + Y_m \underline{u}_m + \cdots + Y_n \underline{u}_n$$

$$\underline{\tilde{X}} = Y_1 \underline{u}_1 + Y_2 \underline{u}_2 + \cdots + Y_m \underline{u}_m$$

$$= [u_1 \; u_2 \; \cdots \; u_m] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}$$

$$m \to n \Rightarrow \tilde{X} \to X$$

- خصوصیات بردارهای پایه (u_i)

orthonormal

$$\underline{u}_i^T \underline{u}_j = \begin{cases} 1 & , i=j \\ 0 & , i \neq j \end{cases}$$

$$\tilde{X} = [u]_m \underline{Y}$$

$$[\underline{u}]_m = [\underline{u}_1 \; \underline{u}_2 \; \cdots \; \underline{u}_m]$$

to find Y_i

$$u_i^T \underline{X} = Y_1 \underbrace{\underline{u}_i^T \underline{u}_1}_{0} + \cdots + Y_i \underbrace{\underline{u}_i^T \underline{u}_i}_{1} + \cdots + Y_m \underbrace{\underline{u}_i^T u_m}_{0}$$

$$Y_i = \underline{u}_i^T \underline{X}$$

- در حالت کلی

$$\underline{Y} = [u]_m^T \underline{\tilde{X}}$$

$\binom{n}{m}$ • حال بجای زیرمجموعه $[U]_m$ را می‌خواهیم نیاز به شاخص داریم
این شاخص، شاخص خطا خواهد بود:

$$\underline{\varepsilon} = \underline{X} - \underline{\tilde{X}} = \sum_{j=m+1}^{n} Y_j \, \underline{u}_j$$

MSE

$$\mathcal{E} = E\{\|\underline{\varepsilon}\|^2\} = E\{\|\underline{X} - \underline{\tilde{X}}\|^2\} = E\{\underline{\varepsilon}^T \underline{\varepsilon}\}$$

$$\mathcal{E} = \sum_{j=m+1}^{n} E\{Y_j^2\}$$

$$Y_j^2 = Y_j \cdot Y_j = (\underline{u}_j^T x)(\underline{u}_j^T x)$$

$$E\{Y_j^2\} = \underline{u}_j^T E\{xx^T\} \cdot \underline{u}_j$$

$$= \underline{u}_j^T [R] \, \underline{u}_j$$

$\hookrightarrow$ correlation matrix of $\underline{X}$

$$\mathcal{E} = \sum_{j=m+1}^{n} \underline{u}_j^T [R] \, \underline{u}_j$$

• پس مسئله تبدیل می‌شود به بهینه‌سازی

$$\mathcal{E} = \sum_{j=m+1}^{n} \underline{u}_j^T [R] \, \underline{u}_j$$

با فرض

$$\underline{u}_i^T \underline{u}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$$\mathcal{E}' = \sum_{j=m+1}^{n} \underline{u}_j^T [R] \, \underline{u}_j + \sum_{j=m+1}^{n} \underbrace{\lambda_i}_{} (1 - \underline{u}_j^T \underline{u}_j)$$

$\hookrightarrow$ Lagrange Multipliers

Subject:
Year. Month. Date. ( )

$$\frac{\partial \varepsilon'}{\partial \underline{u}_j} = 2\left([R]\,\underline{u}_j - \lambda_i\,\underline{u}_j\right) = 0$$

$$\boxed{[R]\,\underline{u}_j = \lambda_j\,\underline{u}_j}$$

→ eigen vector & eigen value

پس برای کمینه کردن ε' ، u_i ها باید بردارهای ویژهٔ $[R]$ باشند.

- اگر

$$\tilde{X}_m = Y_1 \cdot \underline{u}_1 + \cdots + Y_1 \cdot \underline{u}_m + \sum E\{Y_i\}\,\underline{u}_j$$

$$[\Sigma]\,\underline{u}_j = \lambda_j \cdot \underline{u}_j$$

([illegible] notation)

-

$[R]_{n \times n} \longrightarrow \lambda_i$ تا n

می‌خواهیم m تا λ_i را نگه داریم

$\varepsilon = \sum \lambda_i$

پس بزرگترین λ_i را نگه می‌داریم

حال به کمک λ_i های باقی مانده u_i ها را می‌یابیم

- این تبدیل به KL - transformation معروف است.

Karhunen - Loeve Transformation

$$\underline{Y} = [u]^T\,\underline{X}$$

$[u] = [u_1 \ \ldots \ u_m] = [$eigen vectors of $[R]]$

$\underline{X}_{n\times 1}$ *

$\underline{X} = Y_1 \underline{u}_1 + \dots + Y_n \underline{u}_n$

$\tilde{\underline{X}} = Y_1 \underline{u}_1 + \dots + Y_m \underline{u}_m$ $m < n$

$\underline{\varepsilon} = \underline{X} - \tilde{\underline{X}}$

$\varepsilon = \sum_{i=m+1}^{n} [u]^T [R] [u]$

$$\underline{u}_i \underline{u}_j = \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases}$$

$[u] = [\underline{\Phi}_1 \ \underline{\Phi}_2 \ \cdots \ \underline{\Phi}_m]$ ~ KL

* کاربرد در Pattern Recognition

$\omega_1, \dots, \omega_M$

$\underline{X}^i$ $i = 1, \dots, M$

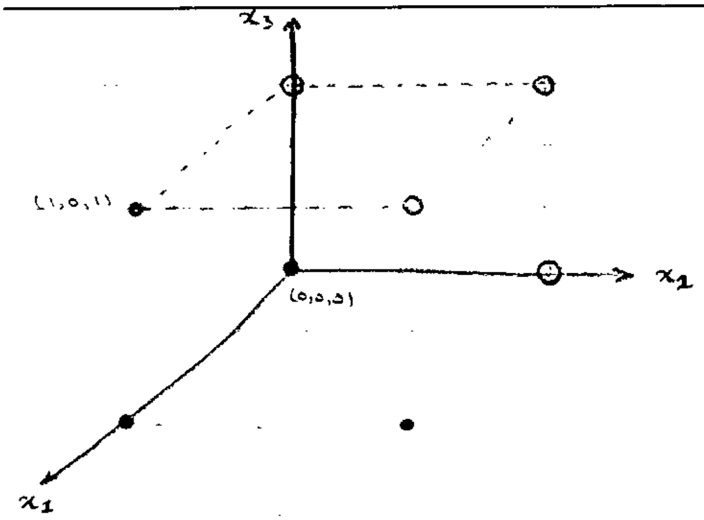
$\underline{X}^i \ \sum_{j=1}^{m} Y_j^i \cdot \underline{\Phi}_j$

• مراحل

1. Find correlation matrix [R]

$[R] = \sum_{i=1}^{M} P(\omega_i) \cdot E\{\underline{X}_i \underline{X}_i^T\}$

2. Obtain the eigenvalues & eigenvectors of [R]
3. $[A] = [\underline{\Phi}_1 \ \dots \ \underline{\Phi}_m]$
4. Transformation



● $\in \omega_1$

○ $\in \omega_2$

	ω_1	ω_2	
$\underline{x}_{11}$ =	$(0,0,0)^T$	$(0,0,1)^T$	= $\underline{x}_{21}$
$\underline{x}_{12}$ =	$(1,0,0)^T$	$(0,1,0)^T$	= $\underline{x}_{22}$
$\underline{x}_{13}$ =	$(1,0,1)^T$	$(0,1,1)^T$	= $\underline{x}_{23}$
$\underline{x}_{14}$ =	$(1,1,0)^T$	$(1,1,1)^T$	= $\underline{x}_{24}$

$$x_{ij} \longrightarrow i = \text{class } \#$$
$$\searrow j = \text{sample } \#$$

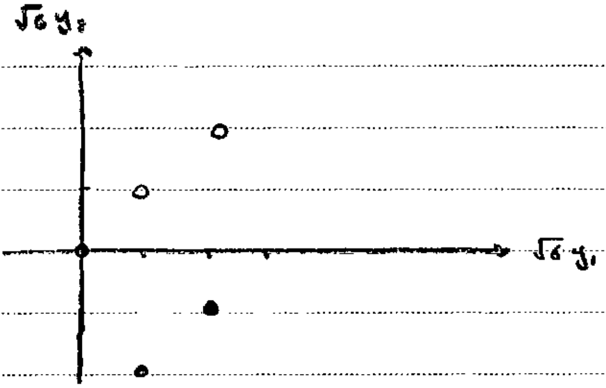
$$[R] = \sum_{i=1}^{2} P_i \, E\{X_i X_i^T\}$$
$$= \frac{1}{2}[R_1] + \frac{1}{2}[R_2]$$
$$= \frac{1}{8}\left(\sum_{j=1}^{4} x_{1j} x_{1j}^T + \sum_{j=1}^{4} x_{2j} x_{2j}^T\right)$$
$$= \frac{1}{4}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$|[R] - \lambda[I]| = 0 \longrightarrow \lambda_1 = 1 \longrightarrow \phi_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$$
$$\searrow \lambda_2 = 1/4 \longrightarrow \phi_2 = \frac{1}{\sqrt{6}}\begin{pmatrix}-2\\1\\1\end{pmatrix}$$
$$\searrow \lambda_3 = 1/4 \longrightarrow \phi_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\-1\end{pmatrix}$$

حال از میان λ_i ، ۲ انتخاب می‌کنیم.

$$[\Phi]_{3\times 2} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \underline{\Phi_1} & \underline{\Phi_2} \end{bmatrix} \Rightarrow \underset{2\times 1}{Y} = \underset{2\times 3}{[\phi]^T} \, \underset{3\times 1}{\underline{X}}$$

ω_1	ω_2
$(0,0)^T$	$\frac{1}{\sqrt{6}}(\sqrt{2},1)^T$
$\frac{1}{\sqrt{6}}(\sqrt{2},-2)^T$	$\frac{1}{\sqrt{6}}(\sqrt{2},1)^T$
$\frac{1}{\sqrt{6}}(2\sqrt{2},-1)^T$	$\frac{1}{\sqrt{6}}(2\sqrt{2},2)^T$
$\frac{1}{\sqrt{6}}(2\sqrt{2},-1)^T$	$\frac{1}{\sqrt{6}}(3\sqrt{2},0)^T$



$MSE = \frac{1}{4}$; خطای MSE برابر eigenvalue است که کنار گذاشتیم

حال اگر بخواهیم به فضای یک بعدی برویم.

1-D

$\lambda_1 = 1$ $\longrightarrow MSE = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$\Phi_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$

$[u] = [\Phi_1]$

$Y_{11} = 0 \longrightarrow Y_{11} = \Phi_1^T \underline{x}_{11} = [\frac{1}{\sqrt{3}} \ \frac{1}{\sqrt{3}} \ \frac{1}{\sqrt{3}}]\begin{bmatrix}0\\0\\0\end{bmatrix} = 0$

$Y_{12} = \frac{1}{\sqrt{3}}$

* کتاب

$$\underline{X} = \sum_{i=1}^{n} Y_i \Phi_i = [\Phi]\underline{Y}$$

$$\hat{\underline{X}}(m) = \sum_{i=1}^{m} Y_i \underline{\Phi}_i + \sum_{i=m+1}^{n} b_i \underline{\Phi}_i$$ ← ثابت

$$\underline{\Delta}(m) = \underline{X} - \hat{\underline{X}}(m)$$
$$= \sum_{i=m+1}^{n} (Y_i - b_i)\underline{\Phi}_i$$

$$\bar{\varepsilon}^2 = \sum_{i=m+1}^{n} E\{(Y_i - b_i)^2\}$$

$$\frac{\partial \bar{\varepsilon}^2}{\partial b_i} = 0 \longrightarrow b_i^* = E\{Y_i\} = \Phi_i^T E\{\underline{X}\}$$

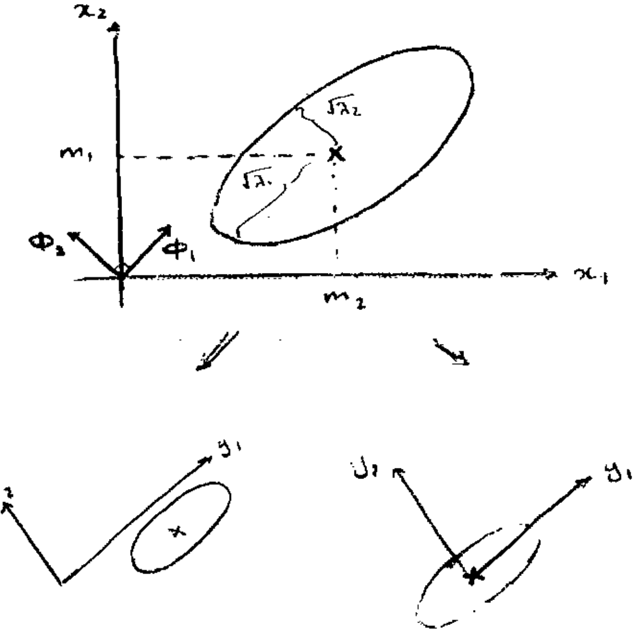
b_i را در $\bar{\varepsilon}^2$ جایگذاری می‌کنیم ...

$$\bar{\varepsilon}^2 = \sum_{i=m+1}^{n} \underline{\Phi}_i^T [\Sigma]_x \Phi_i$$

- مقایسه : به دلیل نداشتن b_i ، بدون میانگین به جای $[\Sigma]_x$ مقدار $[R]$ بدست آمده است.
-

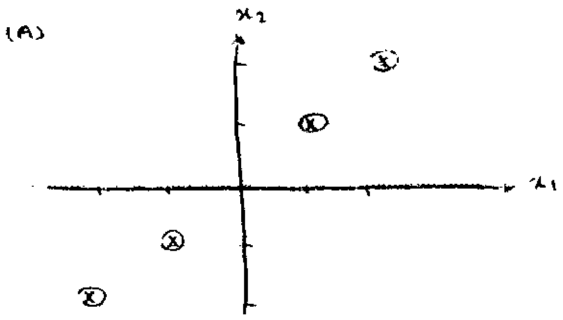
$\text{mean}\{\underline{X}\} = 0 \longrightarrow [\Sigma]_x = [R]$

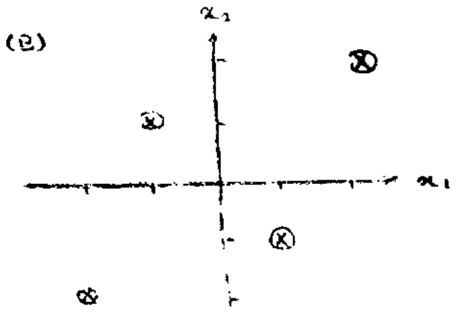
$\text{mean}\{\underline{X}\} \neq 0 \longrightarrow [\Sigma]_x$ better ✓



$$\Phi_i \cdot \Phi_j = \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases}$$

Φ_1 ← به خاطر وجود λ_1 است
در جهت بزرگترین تغییرات
است.





☑

$$(A): [\Sigma] = [R] = \frac{1}{4} \sum_{i=1}^{4} \underline{X}_i \underline{X}_i^T = \begin{bmatrix} \frac{10}{4} & \frac{6}{4} \\ \frac{10}{4} & \frac{10}{4} \end{bmatrix}$$

$$(B): [\Sigma] = \begin{bmatrix} \frac{10}{4} & \frac{6}{4} \\ \frac{6}{4} & \frac{10}{4} \end{bmatrix}$$

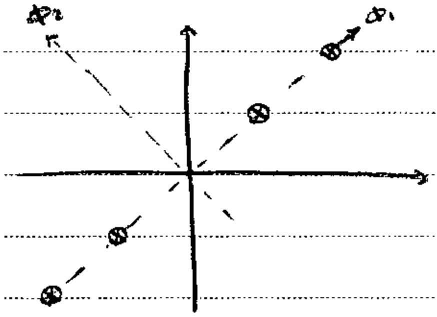
1-D

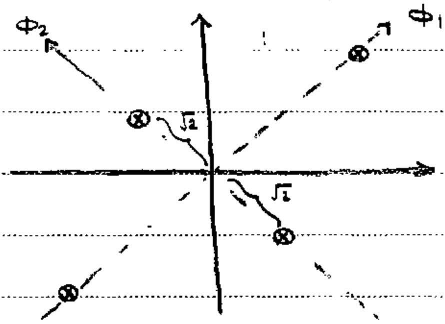
(A) $\lambda_1 = 5$ $\quad \underline{\Phi}_1 = (\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}})^T$

$\lambda_2 = 0$ $\quad \underline{\Phi}_2 = (\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}})^T$ $\quad \Rightarrow \varepsilon = 0$

(B) $\lambda_1 = 4$ $\quad \underline{\Phi}_1 = (\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}})^T$

$\lambda_2 = 1$ $\quad \underline{\Phi}_2 = (\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}})^T$ $\quad \Rightarrow \varepsilon = 1 = \frac{1}{4}(\sqrt{2}^2 + \sqrt{2}^2)$





☒ فشرده سازی با تبدیل KL

بیشترین انرژی در کمترین ضرایب وجود دارد.

البته به Data ← چون به [Σ] نیاز دارد،

به همین دلیل از آن استفاده نمی‌شود. (☒ JPEG ، MPEG)

بجای آن از تبدیلهای نزدیک به بهینه مانند تبدیل کسینوسی استفاده می‌شود)

$$\varepsilon = \sum_{i=m+1}^{n} \lambda_i$$

بجای اینکه بدانیم λ_i های که کنار می‌گذاریم چه نسبتی از کل اطلاعات را در خود دارد، آنها را نرمال می‌کنند. با اینها میزان اطلاعات از دست رفته مشخص می‌شود.

$$\mu_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_i}$$

$$[\Sigma]_y = [\Phi]^T [\Sigma]_x [\Phi] = [\lambda]$$

ماتریس قطری

$$tr([\Sigma]_y) = tr([\Phi]^T [\Sigma]_x [\Phi])$$
$$= tr([\Sigma]_x [\Phi][\Phi]^T)$$
$$= tr([\Sigma]_x [I])$$
$$= tr([\Sigma]_x)$$

پس رابطه زیر نیز برای μ بدست می آید:

$$\mu_i = \frac{\lambda_i}{tr[\Sigma]_x}$$

* شاخص پراکندگی (Scatter Measure)

$$\bar{d}_x^2 = E\{(\text{between sample distance})\}$$
$$= E\{\|\underline{X}_i - \underline{X}_j\|^2\}$$
$$= E\{(\underline{X}_i - \underline{X}_j)^T(\underline{X}_i - \underline{X}_j)\}$$
$$= E\{\underline{X}_i^T \underline{X}_i + \underline{X}_j^T \underline{X}_j\} - E\{X_i^T X_j + X_j^T X_i\}$$

- تشریح داده ها:

identical distribution
independant

$\Big\}$ iid

$$\bar{d}_x^2 = 2E\{\underline{X}^T\underline{X}\} - 2E\{\underline{X}^T\}E\{\underline{X}\}$$
$$= 2\,tr\,[E\{\underline{X}\underline{X}^T\} - m\underline{m}^T]$$
$$= 2\,tr\,[[R] - \underline{m}\,\underline{m}^T]$$
$$= 2\,tr\,[\Sigma]_x$$

Subject:
Year: Month. Date. 70

$\underline{Y} = [\Phi]_m^T \underline{X}$

$[\Phi]_m = [\underline{\phi} \;\; \underline{\phi} \; \cdots \; \underline{\phi}_m]$

$[\Phi]^T [\Phi] = [I]$ ⟶ orthonormal transformation

$\bar{d}_y^2 = 2\, tr\, [\Sigma]_y$ ⟵ distance preserved

$$\bar{d}_y^2 = 2\, tr\, [\Sigma]_y$$
$$= 2\, tr\, ([\Phi]^T [\Sigma]_x [\Phi])$$
$$= 2 \sum_{i=1}^{m} \phi_i^T [\Sigma]_x \, \phi_i$$

• برای Representation

maximize $\bar{d}_y^2$

subject to $\phi_i \phi_j = \begin{cases} 1 , & i = j \\ 0 , & i \neq j \end{cases}$

$$\bar{\varepsilon}^2 = \sum_{i=m+1}^{n} \phi_i^T [\Sigma]_x \, \phi_i$$

⟶ کنار گذاشته ⟶ minimize

$$\bar{d}^2 = \underset{\downarrow}{2} \sum_{i=1}^{m} \phi_i^T [\Sigma]_x \, \phi_i$$

⟶ داده‌های اصلی ⟶ maximize

⟶ best representation!

ارزش داده‌های حاضر دو برابر شده است.

تمرین فصل ۹ : موعد: یکشنبه ۱۰/۱۵

Computer Project : 1 , 2

Problems : 1 , 2 , 3

تعیین موضوع تحقیقاتی : عنوان پروژه + مراجع + کاربردها)

Pattern Recognition - Pattern Recognition Letter

Intl. J. of PR - IEEE PAMI

تحویل پروژه : ۳ هفته بعد از آخرین امتحان دانشکده

• یک شاخص دیگر (population entropy)
شاخصی برای اندازه‌گیری عدم قطعیت در داده‌ها.

$$h_{\underline{x}} = -E\{\ln P_x(\underline{x})\}$$

برای بازنمائی داده‌ها، هدف بیشینه کردن عدم قطعیت داده‌هاست (برخلاف classification)
وابسته به تابع توزیع است.

$$P_x(\underline{x}) = N(\underline{m}_x, [\Sigma_x])$$

$$h_x = E\left\{\frac{1}{2}(\underline{X}-\underline{m}_x)^T[\Sigma]_x^{-1}(\underline{X}-\underline{m}_x) + \frac{1}{2}\ln|[\Sigma]_x| + \frac{n}{2}\ln(2\pi)\right\}$$

$$= \frac{1}{2}E\{tr([\Sigma]_x^{-1}(\underline{X}-\underline{m}_x)(\underline{X}-\underline{m}_x)^T\} + \cdots$$

$$= \frac{1}{2}E\{tr([\Sigma]_x^{-1}[\Sigma]_x) + \cdots$$

$$= \frac{n}{2} + \frac{1}{2}\ln|[\Sigma]_x| + \frac{n}{2}\ln(2\pi)$$

$$h_y = \frac{m}{2} + \frac{1}{2}\ln|[\Sigma]_x| + \frac{m}{2}\ln(2\pi)$$

$$= \frac{m}{2} + \frac{1}{2}\ln|[\Phi]_m^T[\Sigma]_x[\Phi]| + \frac{m}{2}\ln(2\pi)$$

این رابطه باید بیشینه شود با توجه به محدودیت orthonormal trans. :

$$\Sigma_y = [\Phi]_m^T\Sigma_x[\Phi]_m = [\lambda] \longrightarrow \ln\Sigma_y = \ln|[\lambda]| = \sum\ln(\lambda_i)$$

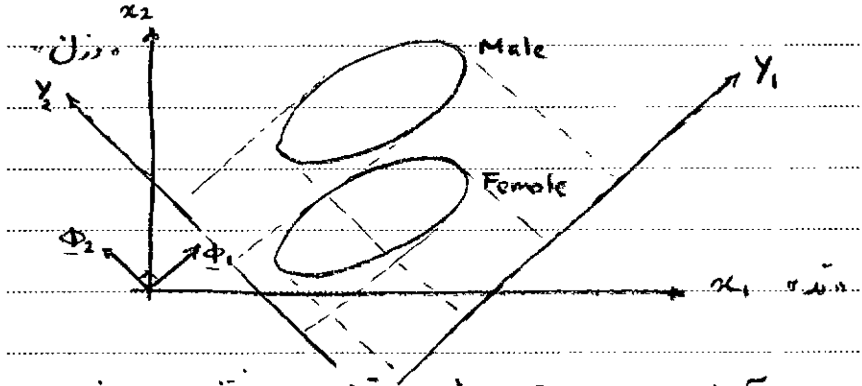
برای بیشینه شدن آن بزرگترین eigen value ها را انتخاب کرده و از eigen vector های متناظر آن‌ها استفاده می‌کنیم.

✱ فصل ۱۰) Feature Extraction & Linear Mapping 4 Classification

- Class Seperability

این شاخص مد نظر است.

به توزیع داده بسته است و به نوع classifier استفاده شده نیز بستگی دارد.



برای کاهش بعد به یک بعد، آن بعد راستای بیشترین تغییرات انتخاب می‌شود. این انتخاب برای representation است.

برای داشتن کمترین خطا برای classification راستای Φ_2 مناسب‌تر است.

- هدف: پیدا کردن linear transformation مناسب به طوری که seperabable بودن را حفظ کند.

$$\underset{m\times1}{\underline{Y}} = \underset{n\times m}{[a]}^T \underset{n\times1}{\underline{X}}$$

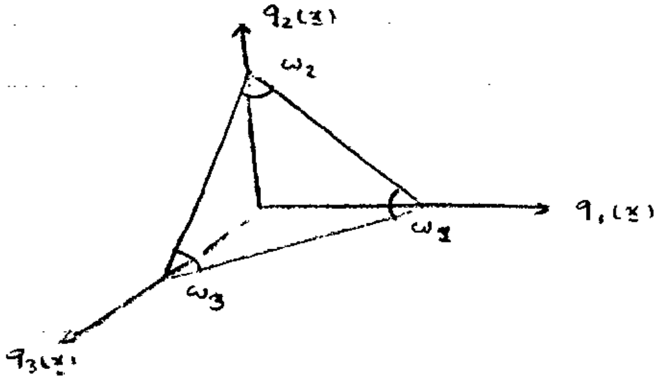
☑ Bayes Classifier

$q_i(\underline{x})$ $i = 1, \dots, M$

$g_i = q_i(\underline{x})$

$$q_i(\underline{x}) \underset{\omega_j}{\overset{\omega_i}{\gtrless}} q_j(\underline{x}) \longrightarrow y_i \gtrless y_j$$

$$\sum_{i=1}^{M} q_i(\underline{x}) = 1$$



M-1 تابع مستقل وجود دارد.

• دو نوع تابع معیار در این فصل استفاده می‌شود.

1. Based on family of scatter matrices
2. Based on bounds

• Scatter Matrices

* Within-Class Scatter Matrix

$$[S]_w = \sum_{i=1}^{M} P_i \,.\, E\{(\underline{X}-\underline{m}_i)(\underline{X}-\underline{m}_i)^T | \omega_i\}$$

$$= \sum_{i=1}^{M} P_i \,.\, [\Sigma]_i$$

میانگین پراکندگی بین کلاسی = میانگین ماتریس‌های پراکندگی تمام M کلاس

* Between-Class Scatter Matrix

$$[S]_b = \sum_{i=1}^{M} P_i (m_i - m_o)(m_i - m_o)^T$$

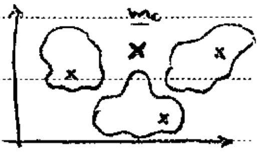
$m_i = E\{X \mid \omega_i\}$

$m_o = \sum_{i=1}^{M} P_i \cdot \underline{m_i} = E\{\underline{X}\}$

* Mixture scatter matrix

$$[S]_m = E\left\{(\underline{X}-\underline{m}_o)(\underline{X}-\underline{m}_o)^T\right\}$$

$$= [S]_w + [S]_b$$



* General Form of Scatter Criteria:

◎ $J_1 = tr\left([S]_2^{-1}[S]_1\right)$

$[S]_2 = [S]_w$ ☑
$[S]_1 = [S]_b$
$J_1 \uparrow \Rightarrow$ seperability $\uparrow$

☑
$[S]_2 = [S]_m$
$[S]_1 = [S]_w$
$J_1 \uparrow \Rightarrow$ seperability $\downarrow$

◎ $J_2 = \ln\left|[S]_2^{-1}[S]_1\right| = \ln|[S]_1| - \ln|[S]_2|$

, $[S]_2 \neq [S]_b \longrightarrow$ چون عکس ندارد

◎ $J_3 = tr\,[S]_1 - \mu\,(tr\,[S]_2 - c)$

@ $J_4 = \dfrac{tr\ [S]_1}{tr\ [S]_2}$

Optimum Linear Transform.

$\underline{Y} = [a]^T \underline{X}$

there is no constraint on [a] to be orthonormal $([a]^T[a] \neq [I])$

$[\Sigma]_Y = [a]^T [\Sigma]_X [a]$

$[S]_Y = [a]^T [S]_X [a]$ $\rightsquigarrow$ scatter matrix

$$J_1(m) = tr\ ([S]^{-1}_{2Y} [S]_{1Y})$$
$$= tr\ \{([a]^T [S]_{2X} [a])^{-1} ([a]^T [S]_{1X} [a])\}$$

$\dfrac{d J_1}{d [a]} = 0$ $\longrightarrow$ Appendix A · A.15

$$\frac{\partial J_2(m)}{\partial [a]} = -2\ \underline{S}_{2X} \underline{A} \underline{S}_{2Y}^{-1} \underline{S}_{1Y} \underline{S}_{2Y}^{-1} + 2\ \underline{S}_{1X} \underline{A} \underline{S}_{2Y}^{-1} = \underline{0}$$

$\Rightarrow S_{1X} A S_{2Y}^{-1} = S_{2X} A S_{2Y}^{-1} S_{1Y} S_{2Y}^{-1}$

$\Rightarrow S_{2X}^{-1} S_{1X} A S_{2Y}^{-1} = A S_{2Y}^{-1} S_{1Y} S_{2Y}^{-1}$

$\Rightarrow (S_{2X}^{-1} S_{1X}) A = A (S_{2Y}^{-1} S_{1Y})$

$\underline{Y} = A^T \underline{X}$

$\underline{Z} = B^T \underline{Y}$ linear transformation

$B^T S_{1Y} B = [\mu]_m$ – [illegible]

$B^T S_{2Y} B = [I]_m$

در واقع S_{1Y} ، S_{2Y} را به طور همزمان قطری می کنیم.

$$
\begin{aligned}
tr\left(S_{2Z}^{-1} S_{1Z}\right) &= tr\left((B^T S_{2Y} B)^{-1} (B^T S_{1Y} B)\right) \\
&= tr\left(B^{-1} S_{2Y} B^{-T} B^T S_{1Y} B\right) \\
&= tr\left(S_{2Y}^{-1} S_{1Y} B B^{-1}\right) \\
&= tr\left(S_{2Y}^{-1} S_{1Y}\right)
\end{aligned}
$$

در این تبدیل ، مقدار criteria تغییر نمی کند.

$$
\begin{aligned}
B^T S_{1Y} B &= [\mu] \Rightarrow \\
B^T S_{1Y} &= [\mu] B^{-1} \Rightarrow \\
S_{1Y} &= B^{-T} [\mu] B^{-1}
\end{aligned}
$$

و

$$
\begin{aligned}
B^T S_{2Y} B &= [I] \Rightarrow \\
B^T S_{2Y} &= B^{-1} \Rightarrow \\
S_{2Y} &= B^{-T} B^{-1}
\end{aligned}
$$

با جایگذاری

$$
\begin{aligned}
S_{2Y}^{-1} S_{1Y} &= (B^{-T} B^{-1})^{-1} (B^{-T} [\mu] B^{-1}) \\
&= (B B^T)(B^{-T} [\mu] B^{-1}) \\
&= B [\mu] B^{-1}
\end{aligned}
$$

پس

$$
\begin{aligned}
(S_{2X}^{-1} S_{1X}) A &= A (S_{2Y}^{-1} S_{1Y}) \Rightarrow \\
(S_{2X}^{-1} S_{1X}) A &= A (B [\mu] B^{-1}) \Rightarrow \\
(S_{2X}^{-1} S_{1X}) AB &= AB ([\mu])
\end{aligned}
$$

AB ما را از فضای X به Z می برد.

ضمناً $[\mu]$ eigen value $(S_{2Y}^{-1} S_{1Y})$ است لذا eigen value $(S_{2X}^{-1} S_{1X})$ نیز هست.

Subject:

Year. Month. Date. ( )

با توجه به قطری کردن همزمان، مقادیر ویژه ماتریس [μ] اعداد حقیقی و مثبت هستند.

$$J(n) = tr\,(S_{2x}^{-1}\,S_{1x}) = \lambda_1 + \dots + \lambda_n$$
$$J(m) = tr\,(S_{2y}^{-1}\,S_{1y}) = \mu_1 + \dots + \mu_m$$

برای بیشینه یا کمینه کردن J به ترتیب بزرگترین یا کوچکترین مقدارهای ویژه استفاده می‌شود

☑

2- class

$$J_1 = tr\,([S]_w^{-1}\,[S]_b) \quad \begin{matrix} \longrightarrow S_2 = [S]_w \\ \searrow S_1 = [S]_b \end{matrix}$$

داده‌ها فشرده $\Rightarrow [S]_w \downarrow \;\Rightarrow\; [S]_w^{-1}\uparrow$

فاصله داده‌ها زیاد $\Rightarrow [S]_b \uparrow$

$\mapsto J \sim$ maximize

$$\underline{Y} = [A]^T \underline{X}$$

$P_i =$ احتمال اولیه کلاس i

$$[S]_b = \sum_{i=1}^{M} P_i\,(\underline{m}_i - m_0)(\underline{m}_i - m_0)^T \quad \longleftarrow m_0 = P_1\underline{m}_1 + P_2\underline{m}_2$$

$$= P_1(m_1 - m_0)(m_1 - m_0)^T + P_2(m_2 - m_0)(m_2 - m_0)^T$$

$$= P_1(m_1 - (\sim))(m_1 - (\sim))^T + P_2(m_2 - (\sim))(m_2 - (\sim))^T$$

$$= \dots$$

$$= [P_1 P_2^2 + P_1^2 P_2]\,(m_2 - m_1)(m_2 - m_1)^T$$

$$= P_1 P_2\,(P_1 + P_2)(m_2 - m_1)(m_2 - m_1)^T$$

$$= P_1 P_2\,(m_2 - m_1)(m_2 - m_1)^T$$

تنها یک سطر مستقل خطی داریم، یک مقدار ویژه مثبت می‌آید پس:

$\text{Rank}([S]_b) = 1$

$\text{Rank}([S]_w^{-1}) =$ full rank

$\text{Rank}([S]_w^{-1}[S_b]) = 1 \sim$ یک بعد مستقل

*پس تنها یک سطر مستقل دارد:

$$tr(S_w^{-1}S_b) = \lambda_1$$

$$\lambda_1 = \text{eigen value}\ (S_w^{-1}S_b) \neq 0$$

$$= tr\left(S_w^{-1}(P_1P_2(m_2-m_1)(m_2-m_1)^T\right)$$

$$= P_1P_2\left(tr((m_2-m_1)^T S_w^{-1}(m_2-m_1))\right)$$

$$= P_1P_2\left((m_2-m_1)^T S_w^{-1}(m_2-m_1)\right)$$

$$\underline{\Phi}_1 = \frac{[S]_w^{-1}(\underline{m}_2-\underline{m}_1)}{\|[S]_w^{-1}(m_2-m_1)\|} \quad *$$

$$(S_w^{-1}S_b)\,\underline{\Phi}_1 = \lambda_1\,\underline{\Phi}_1$$

*پس اگر نیاز به تبدیل باشد:

$$\underline{Y} = [a]^T\underline{x}$$

$$[a] = \underline{\Phi}_1$$

$$Y = \underline{\Phi}_1\,\underline{X} = cte$$

بیان دو بعد به یک بعد:

$$Y = c\,[[S]_w^{-1}(m_2-m_1)]^T\underline{X}$$

$\hookrightarrow$ متناسب

$$\boxed{Y = c\,(m_2-m_1)^T[S]_w^{-1}\underline{X}}$$

*مثال به Bayes Classifier

$$g_i(\underline{x}) \qquad i=1,\dots,M$$

$$g_i \underset{\omega_j}{\overset{\omega_i}{\gtrless}} g_z$$

$$g_1+\dots+g_M=1$$

در اینجا به وسیله تابع q فضا به 1D برده می‌شود.

* مثالی با Linear classification

$h(\underline{x}) = \underline{v}^T \underline{x}$

$\underline{v} = (\underbrace{S \Sigma_1 + (1-S) \Sigma_2}_{})^{-1} (m_2 - m_1)$

کاری که $[S]_w$ انجام می‌دهد → به نوعی پراکندگی Σ ها را نشان می‌دهد

M - class ☑

$[S]_b = \sum_{i=1}^{M} P_i \quad (m_i - m_0)(m_i - m_0)^T$

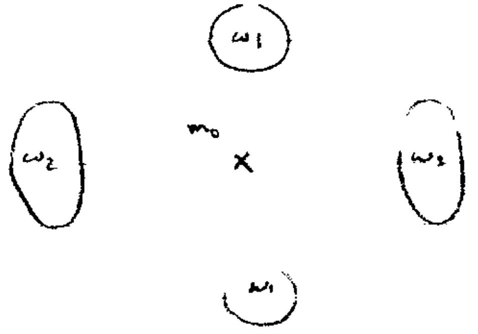
$J = tr([S]_w^{-1} [S]_b)$

$Rank([S]_b) = M-1$

$Rank(J) = M-1 \longrightarrow \lambda_1, \dots, \lambda_{M-1} \neq 0 \longrightarrow \underline{\Phi}_1, \dots, \underline{\Phi}_{M-1}$

☑ حالت خاص

$m_0 = m_1 = m_2$



این شاخص خوب عمل می کند. چون پراکندگی حول میانگین را نشان می دهد.

☑

$S_2 = S_m$

$S_1 = S_w$

$J = tr(S_m^{-1} S_w)$

باید ببینید ،،

$$(S_{2x}^{-1} S_{1x})\, AB = AB\,([\mu]) \Rightarrow$$

$$(S_m^{-1} S_\omega)[\phi] = [\phi][\lambda] \Rightarrow$$

$$\begin{cases} \text{eigen value matrix } S_m^{-1} S_w = [\lambda] \\ \text{eigen vector matrix } S_m^{-1} S_w = [\phi] \end{cases}$$

$$\begin{cases} S_\omega [\phi] = S_m [\phi][\lambda] \Rightarrow \\ S_m = S_\omega + S_b \end{cases}$$

$$S_\omega [\phi] = (S_\omega + S_b)[\phi][\lambda] \Rightarrow$$

$$S_b [\phi][\lambda] = S_\omega [\phi]([I] - [\lambda]) \Rightarrow$$

$$S_\omega^{-1} S_b [\phi][\lambda] = [\phi](\underline{I} - \underline{\lambda}) \Rightarrow$$

$$S_\omega^{-1} S_b [\phi] = [\phi]([\lambda]^{-1} - [I])$$

اگر مقادیر ویژه ی $S_m^{-1} S$ را مرتب کنیم

$$\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \dots \geqslant \lambda_n$$

حال مقادیر ویژه $S_\omega^{-1} S_b$ را مرتب می‌کنیم.

$$\left(\frac{1}{\lambda_1} - 1\right) \leqslant \left(\frac{1}{\lambda_2} - 1\right) \leqslant \dots \leqslant \left(\frac{1}{\lambda_n} - 1\right)$$

Optimization of $J_2(m)$ *

$$J_2(m) = \ln |S_{1y}| - \ln |S_{2y}|$$

$$\underline{Y} = [A]^T \underline{x}$$

$$J_2(m) = \ln |A^T S_{1x} A| - \ln |A^T S_{2x} A|$$

$$\frac{\partial J_2(m)}{\partial A} = [0] \longrightarrow \text{Appendix A : A.28}$$

$$\frac{\partial J}{\partial A} = 2\left\{ S_{1X} A S_{1Y}^{-1} - S_{2X} A S_{2Y}^{-1} \right\} = 0$$

$$S_{1X} A S_{1Y}^{-1} = S_{2X} A S_{2Y}^{-1}$$

$$S_{2X}^{-1} S_{1X} A S_{1Y} S_{1Y}^{-1} = S_{2X}^{-1} S_{2X} A S_{2Y}^{-1} S_{1Y}$$

$$(S_{2X}^{-1} S_{1X}) A = A (S_{2Y}^{-1} S_{1Y})$$

مشابه J_1 شد.

$$J_2(m) = \ln \frac{|S_{1Y}|}{|S_{2Y}|} = \ln |S_{2Y}^{-1} S_{1Y}|$$

$$= \ln \lambda_1 + \dots + \ln \lambda_m$$

نقصان ! J_1 است که S_{2Y} غیر وارون پذیر باشد.

Rank $(S_b) \neq$ full rank

$|S_b| = 0$

پس

$$S_2 \neq |S|_b$$

✎

مجموعه داده واقعی :

http:// kdd . ics . uci . edu

✎ موضوعات تحقیقاتی

1. Combination of Classifiers
2. Non_Linear Classifier
3. Clustering
4. SVMs
5. Density Estimation

6. Application
7. Validation
8. Feature Selection & Extraction
9. Micro-array Data Processing

Validation ◎

* چرا؟

• وجود پارامترهای آزاد

☑

KNN → k

Parzen → h

Feature Reduction → m

• نکات مهم:

- انتخاب مدل
- خطا به سمت مقدار واقعی

• تعداد داده محدود

- Naive → استفاده از کل داده‌ها برای آموزش
  - استفاده از همان داده‌ها برای Test
  - ضعف در generalization
  - خطای کمتر از مقدار واقعی (خوشبینانه)
- روشهای دیگر

• روشهای مناسب برای تخمین عملکرد

- hold out → کل داده‌ها: training + test
  - کم شدن تعداد داده‌های را نداشت
  - اصلاح با روشهای Resampling

- Random Sub Sampling

- تعدادی داده به طور تصادفی از درون داده‌ها برداشته شده برای test استفاده می‌شود.
- این کار k بار صورت می‌گیرد.
- خطای حاصل از این k بار را میانگین می‌گیریم.
- برداشتن داده‌ها بدون جایگزینی است.

- k-Fold Cross Validation

- کل داده به k قسمت تقسیم می‌شود.
- هر بار یک بخش را برای test کنار می‌گذاریم.
- خطای حاصل از این k بار را میانگین می‌گیریم.

- Leave One Out

- یک داده کنار گذاشته می‌شود برای Test
- برای n-1 داده مانده، طراحی انجام می‌شود، و با همان یکی که خارج مانده است test می‌شود.
- به ازای همه n نمونه این کار انجام می‌شود.
- خطای حاصل میانگین گرفته است.

- چند Fold ؟

- تعداد زیادتر ← ↓ Bias
  - ↑ واریانس
  - ↑ زمان

✓ k=3 ← Large Datasets ☑

k=N ← ✓ Leave One Out ← Small Datasets

K=10 ← Normal

- Bootstrap

- مانند Random Sub Sampling
- با جایگذاری ← یعنی یک داده می‌تواند چندبار در داده‌های آموزش استفاده شده
داده‌های استفاده نشده در آزمایش استفاده می‌شود.
- واریانس ↑

- Three Way Data Estimate

- داده‌ها ← Training Set
validation Set ← tune کردن پارامترهای classifier
Test Set

Clustering

- هدف :

روش بدون ناظر
دسته‌بندی داده‌ها و برچسب زدن به آنها بدون حضور ناظر

- Sensible
- مراحل کلی

1. Feature Selection
2. Proximity Measure → Similarity
   → Dissimilarity
3. Clustering Criterion → Compact Cluster
   → Elongated Cluster
4. Clustering Alg.
5. Validation → Expert
6. Interpret
7. Tendency

- انواع ویژگیها
  - Nominal → اسمی
  - Ordinal → ترتیبی
  - Time Scale
  - ...

- انواع الگوریتم
  - Sequential
  - Hierarchical → Agglomerative (Bottom - up)
    → Divisive (Top - Down)

Fitness Functions —

Ad-Hoc —

- خوشه‌بندی سلسله مراتبی : روش تجمعی

* Agglomerative Clustering Alg

* هر cluster با یک عنصر آغاز می‌شود — خوشه‌ها ترکیب می‌شوند — در بالاترین سطح یک خوشه می‌ماند

* روش Bottom-Up

* الگوریتم کلی

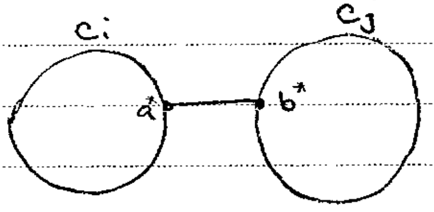
1. Begin with N cluster
   each sample $\in C$

2. Repeat step 3 a total of N-1 times

3. Find the most similar clusters C_i & C_j
   merge C_i & C_j into one cluster

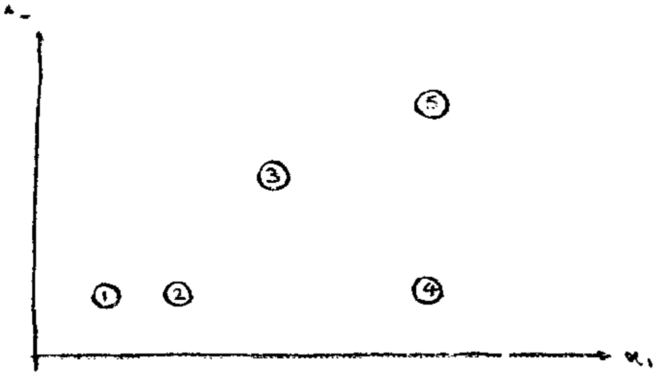
* روش Single-Linkage (SL)
Min Method
Nearest Neighbor Method

$$D_{SL}(C_i, C_j) = \min_{\substack{a \in C_i \\ b \in C_j}} d(a, b)$$



$d(a,b) \equiv$ distance between the samples a, b

Sample #	(x , y)
1	(4 , 4)
2	(8 , 4)
3	(15 , 8)
4	(24 , 4)
5	(24 , 12)



تشکیل ماتریس شباهت نمونه‌ها

(dis) Similarity Matrix

	1	2	3	4	5
1	_	4	11.7	20	21.5
2		_	8.1	16	17.9
3			_	9.8	9.8
4				_	8
5					_

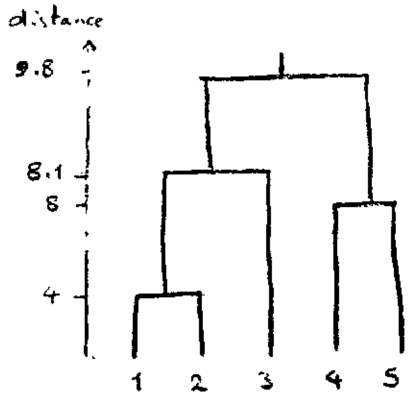
level 1 : {1} {2} {3} {4} {5}

level 2 : {1,2} {3} {4} {5}

level 3 : {1,2} {3} {4 , 5}

level 4 : {{1,2}, 3} {4 , 5}

level 5 : {{1,2}, 3, {4 , 5}}



نمودار Dendrogram ـــ ماتریسهای بجای ما با بردن اطلاعات این منحنی را ذخیره آنها وجود دارد.

این روش Compact cluster بوجود آورد

* روش (Ch) Compelete Linkage

Maximum Method

Farthest Neighbor Method

$$D_{ch}(C_i, C_j) = \max_{\substack{a \in C_i \\ b \in C_j}} d(a,b)$$

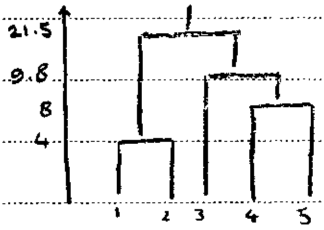
برای ادغام، کمترین فاصله استفاده می‌شود. این D تنها فاصله بین خوشه‌هاست.

Similarity Matrix ⇒

$$\begin{array}{c|cccc} & \{1,2\} & 3 & 4 & 5 \\ \hline \{1,2\} & 0 & 11.7 & 20 & 21.5 \\ 3 & & 0 & 9.8 & 9.8 \\ 4 & & & 0 & \boxed{8.0} \\ 5 & & & & 0 \end{array} \Rightarrow \begin{array}{c|ccc} & \{1,2\} & 3 & \{4,5\} \\ \hline \{1,2\} & 0 & 11.7 & 21.5 \\ 3 & & 0 & \boxed{9.8} \\ \{4,5\} & & & 0 \end{array}$$

$\{\{1,2\},\{3,\{4,5\}\}\}$

elongated cluster



* روش Unweighted Pair-Group Method Using Arithmatic Average (UPGMA)

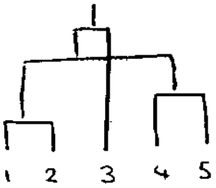
برای از بین بردن حساسیت به نویز ← روشهای قبلی به تک داده حساس بود

$$D(C_i, C_j) = \frac{1}{n_i \times n_j} \sum_{\substack{a \in C_i \\ b \in C_j}} d(a,b)$$

$C_i = i^{th}$ cluster with n_i members

Similarity Matrix $\Rightarrow$

$$\begin{array}{c|cccc} & \{1,2\} & 3 & 4 & 5 \\ \hline \{1,2\} & 0 & 9.9 & 18 & 19.7 \\ 3 & & 0 & 9.5 & 9.8 \\ 4 & & & 0 & \boxed{8.0} \\ 5 & & & & 0 \end{array} \Rightarrow \begin{array}{c|ccc} & \{1,2\} & 3 & \{4,5\} \\ \hline \{1,2\} & 0 & 9.9 & \boxed{8.9} \\ 3 & & 0 & 9.8 \\ \{4,5\} & & & 0 \end{array}$$



$$D(C_1, C_2) = \frac{1}{2\times1}\left(d(1,3) + d(2,3)\right) = 9.9$$

$$D(C_{12}, C_{45}) = \frac{1}{2\times2}\left(d(\{1,2\},4) + d(\{1,2\},5)\right) = 8.9$$

Ward's Method روش *

Min Variance Method

$C_1, C_2, \ldots, C_N$ ~ initial clusters

smallest square error

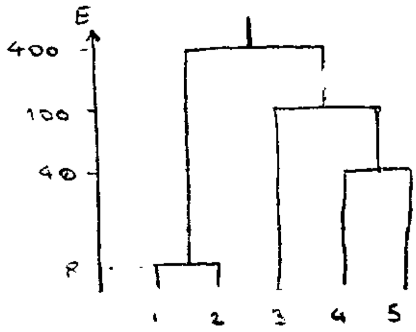
$\underline{x}^1, \ldots, \underline{x}^N$

$$\underline{x}^{(i)} = \begin{bmatrix} x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}$$

$$E = \sum_{i=1}^{N} \sum_{j=1}^{n} (x_j^i - \mu_j)^2$$

$$E = N\sigma^2 = N(\sigma_1^2 + \cdots + \sigma_L^2)$$

clusters	squared error
$\{1,2\}, \{3\}, \{4\}, \{5\}$	$\mu = \frac{1}{2}\left[\binom{4}{4} + \binom{8}{4}\right] = \begin{bmatrix} 6 \\ 4 \end{bmatrix} \longrightarrow E_{1,2} = (4-6)^2 + (8-6)^2 + (4-4)^2 + (4-4)^2 = 8$ $E = 8+0+0+0+0 = \underline{\underline{8}}$ ✓
$\{1,3\}, \{2\}, \{4\}, \{5\}$	68.5
$\{1,4\}, \{2\}, \{3\}, \{5\}$	200
$\{1,5\}, \{2\}, \{3\}, \{4\}$	232
$\{1\}, \{2,3\}, \{4\}, \{5\}$	32
$\{1,2\}, \{3\}, \{4,5\}$	40 ✓



خوشه‌ها براساس اطلاعات پراکندگی تشکیل می‌شود.

Subject:
Year. Month. Date. 80

• خوشه‌بندی سلسله‌مراتبی : روشهای تقسیم کننده.

* روش Forgy

input : K = # of clusters

K x seed

1. initialize the cluster centroid (seed points)
2. each sample
   find the closest cluster (based on their centroids)
   assign sample to its cluster
3. if no samples changed clusters in step 2 → stop
4. compute the centroids of resulting clusters & goto 2

K = 2 ☑

seeds : $\begin{pmatrix}4\\4\end{pmatrix}$, $\begin{pmatrix}8\\4\end{pmatrix}$

	(4, 4)	(8, 4)	(15, 8)	(24, 4)	(24, 12)
Nearest centroid	$\begin{pmatrix}4\\4\end{pmatrix}$	$\begin{pmatrix}8\\4\end{pmatrix}$	$\begin{pmatrix}8\\4\end{pmatrix}$	$\begin{pmatrix}8\\4\end{pmatrix}$	$\begin{pmatrix}8\\4\end{pmatrix}$

$\{\begin{pmatrix}4\\4\end{pmatrix}\}$, $\{\underbrace{\begin{pmatrix}8\\4\end{pmatrix}, \begin{pmatrix}15\\8\end{pmatrix}, \begin{pmatrix}24\\4\end{pmatrix}, \begin{pmatrix}24\\12\end{pmatrix}}\}$

seeds : $\begin{pmatrix}4\\4\end{pmatrix}$, $\begin{pmatrix}17.75\\7\end{pmatrix}$

	$\begin{pmatrix}4\\4\end{pmatrix}$	$\begin{pmatrix}4\\4\end{pmatrix}$	$\begin{pmatrix}17.75\\7\end{pmatrix}$	$\begin{pmatrix}17.75\\7\end{pmatrix}$	$\begin{pmatrix}17.75\\7\end{pmatrix}$

$\{\begin{pmatrix}4\\4\end{pmatrix}, \begin{pmatrix}8\\4\end{pmatrix}\}$, $\{\begin{pmatrix}15\\8\end{pmatrix}, \begin{pmatrix}24\\4\end{pmatrix}, \begin{pmatrix}24\\12\end{pmatrix}\}$

seeds : $\begin{pmatrix}6\\4\end{pmatrix}$, $\begin{pmatrix}21\\8\end{pmatrix}$

	$\begin{pmatrix}6\\4\end{pmatrix}$	$\begin{pmatrix}6\\4\end{pmatrix}$	$\begin{pmatrix}21\\8\end{pmatrix}$	$\begin{pmatrix}21\\8\end{pmatrix}$	$\begin{pmatrix}21\\8\end{pmatrix}$
	C_1	C_1	C_2	C_2	C_2

PAPCO

No Cluster Change ⇒ Stop

* روش k-means

input : K = # of clusters , K seeds

1. Begin with K cluster
   each cluster consists of 1 sample
   for remaing $(N-K)$ sample
      find the closest clusters
   after each sample is assigned
      recompute the centroid
2. Go through data a second time
   DO NOT update the centroids

$K=2$ ☑

seeds: $\begin{pmatrix}8\\4\end{pmatrix}$, $\begin{pmatrix}24\\4\end{pmatrix}$

§ $\underline{x} = \begin{pmatrix}15\\8\end{pmatrix} \rightarrow d(\underline{x}, \bar{c}_1)$ ✓ $\rightarrow C_1 \rightarrow \begin{cases} C_1 : \left\{\begin{pmatrix}8\\4\end{pmatrix}, \begin{pmatrix}15\\8\end{pmatrix}\right\} \rightarrow \bar{c}_1 = \begin{pmatrix}11.5\\6\end{pmatrix} \\ C_2 : \left\{\begin{pmatrix}24\\4\end{pmatrix}\right\} \rightarrow \bar{c}_2 = \begin{pmatrix}24\\4\end{pmatrix}\end{cases}$
$\searrow d(\underline{x}, \bar{c}_2)$

§ $\underline{x} = \begin{pmatrix}4\\4\end{pmatrix} \rightarrow d(\underline{x}, \bar{c}_1)$ ✓ $\rightarrow C_1 \rightarrow \bar{c}_1 = \begin{pmatrix}9\\5.3\end{pmatrix}$
$\searrow d(\underline{x}, \bar{c}_2)$ $\searrow \bar{c}_2 = \begin{pmatrix}24\\4\end{pmatrix}$

§ $\underline{x} = \begin{pmatrix}24\\12\end{pmatrix} \rightarrow d(\underline{x}, \bar{c}_1) \rightarrow C_2 \rightarrow \bar{c}_1 = \begin{pmatrix}9\\5.3\end{pmatrix}$
$\searrow d(\underline{x}, \bar{c}_2)$ ✓ $\searrow \bar{c}_2 = \begin{pmatrix}24\\8\end{pmatrix}$

حال به مرحله دوم می‌رویم با $\bar{c}_1 = \begin{pmatrix}9\\5.3\end{pmatrix}$, $\bar{c}_2 = \begin{pmatrix}24\\8\end{pmatrix}$ که تغییر نمی‌کند.

$C_1 = \left\{ \begin{pmatrix}8\\4\end{pmatrix}, \begin{pmatrix}15\\8\end{pmatrix}, \begin{pmatrix}4\\4\end{pmatrix} \right\}$

$C_2 = \left\{ \begin{pmatrix}24\\4\end{pmatrix}, \begin{pmatrix}24\\12\end{pmatrix} \right\}$