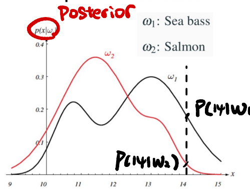
CISC3024 Final Review – CH2. Bayes Theory.

Lecture Note. Example 1.

- Example 1: Sea bass vs. Salmon



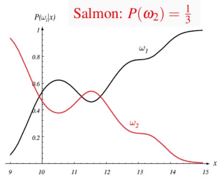
Prior $\begin{cases} P(\omega_1) = \frac{2}{3} \\ P(\omega_2) = \frac{1}{3} \end{cases}$

What will be the posterior probability for either fish type look like?

Likelihood of observing x (lightness) for given class label ω_j (fish type)

16

- Example 1: Sea bass vs. Salmon (Cont.)



Posterior probability for either fish type

For example: $x = 14$

$P(\omega_1|x) = 0.92$
$P(\omega_2|x) = 0.08$

For each value of x
- the higher curve yields the output of Bayesian decision
- the posteriors of either curve sum to 1.0

17

Lecture Note. Example 2

- **Example 2: Cancer detection**

Problem statement

- A new medical test is used to detect whether a patient has a certain cancer or not. His test result is either + (positive) or - (negative)
- For the patient with this cancer, the probability of returning positive test result is 0.98
- For the patient without this cancer, the probability of returning negative test result is 0.97
- The probability of any person having this cancer is 0.008

Question

If positive test result is returned for some person, does the patient have this cancer or not?

Known

$P(+|ca) = 0.98 \Rightarrow P(-|ca) = 0.02$

$P(-|\neg ca) = 0.97 \Rightarrow P(+|\neg ca) = 0.03$

$P(ca) = 0.08, \quad P(\neg ca) = 0.92$

$$P(ca|+) = \frac{P(+|ca)\cdot P(ca)}{P(+)}$$

$$= \frac{P(+|ca)\cdot P(ca)}{P(+|ca)\cdot P(ca) + P(+|\neg ca)\cdot P(\neg ca)}$$

$$= \frac{0.98 \times 0.08}{0.98\times 0.08 + 0.03 \times 0.92}$$

$$= 0.7396$$

$$P(\neg ca|+) = \frac{P(+|\neg ca)\cdot P(\neg ca)}{P(+|\neg ca)\cdot P(\neg ca) + P(+|ca)\cdot P(ca)}$$

$$= \frac{0.03\times 0.92}{0.03\times 0.92 + 0.98\times 0.08}$$

$$= 0.2604.$$

$\therefore$ More likely to have cancer.

Homeworks

2.2 In a two-class one-dimensional problem, the pdfs are the Gaussians $\mathcal{N}(0, \sigma^2)$ ω_1 and $\mathcal{N}(1, \sigma^2)$ ω_2 for the two classes, respectively. Show that the threshold x_0 minimizing the average risk is equal to

$$x_0 = 1/2 - \sigma^2 \ln \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$$

where $\lambda_{11} = \lambda_{22} = 0$ has been assumed.

Posterior Prob. Distribution:

$$P(x|\omega_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \qquad \mu=0,\ \sigma$$

$$P(x|\omega_2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2\sigma^2}} \qquad \mu=1,\ \sigma$$

Cost Matrix:

$\longrightarrow$ Ground Truth

$\downarrow$ Action

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$$

Average Risks:

$$R(\alpha_i|x) = \sum_{j=1}^{N} \lambda(\alpha_i|\omega_j) \cdot P(\omega_j|x)$$

$$= 0 + \lambda_{12} \cdot P(\omega_2|x)$$

$$= \lambda_{12} \cdot P(\omega_2|x)$$

$$R(\alpha_2|x) = \lambda_{21} \cdot P(\omega_1|x) + 0$$

$$= \lambda_{21} \cdot P(\omega_1|x)$$

Minimize Average Risk:

$$R(\alpha_1|x_0) = R(\alpha_2|x_0)$$

$$\Rightarrow \lambda_{12} \cdot P(\omega_2|x_0) = \lambda_{21} \cdot P(\omega_1|x_0)$$

$$\Rightarrow \frac{\lambda_{21}}{\lambda_{12}} = \frac{P(\omega_1|x_0)}{P(\omega_2|x_0)}$$

$$\Rightarrow \frac{\lambda_{21}}{\lambda_{12}} = \frac{P(x_0|\omega_1) \cdot P(\omega_1)}{P(x_0|\omega_2) \cdot P(\omega_2)}$$

$$\Rightarrow \frac{P(x_0|\omega_1)}{P(x_0|\omega_2)} = \frac{\lambda_{12} \cdot P(\omega_1)}{\lambda_{21} \cdot P(\omega_2)}$$

$$\Rightarrow \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_0^2}{2\sigma^2}}}{\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_0-1)^2}{2\sigma^2}}} = \frac{\lambda_{12} \cdot P(\omega_1)}{\lambda_{21} \cdot P(\omega_2)}$$

$$\Rightarrow e^{\frac{(x_0-1)^2 - x_0^2}{2\sigma^2}} = \frac{\lambda_{12} \cdot P(\omega_1)}{\lambda_{21} \cdot P(\omega_2)}$$

$$\Rightarrow \frac{(x_0-1)^2 - x_0^2}{2\sigma^2} = \ln \frac{\lambda_{12} \cdot P(\omega_1)}{\lambda_{21} \cdot P(\omega_2)}$$

$$\Rightarrow x_0^2 - 2x_0 + 1 - x_0^2 = 2\sigma^2 \ln \frac{\lambda_{12} \cdot P(\omega_1)}{\lambda_{21} \cdot P(\omega_2)}$$

$$\Rightarrow 2x_0 - 1 = -2\sigma^2 \ln \frac{\lambda_{12} \cdot P(\omega_1)}{\lambda_{21} \cdot P(\omega_2)}$$

$$\Rightarrow x_0 = \frac{1}{2} - \sigma^2 \ln \frac{\lambda_{12} \cdot P(\omega_1)}{\lambda_{21} \cdot P(\omega_2)}$$

2.5 Consider a two (equiprobable) class, one-dimensional problem with samples distributed according to the Rayleigh pdf in each class, that is,

$$p(x|\omega_i) = \begin{cases} \frac{x}{\sigma_i^2} \exp\left(\frac{-x^2}{2\sigma_i^2}\right) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Compute the decision boundary point $g(x) = 0$. Discriminant Function?

$$g(x) = P(\omega_1|\hat{x}) - P(\omega_2|\hat{x}) = 0$$

$$\Rightarrow P(\omega_1|\hat{x}) = P(\omega_2|\hat{x})$$

$$\Rightarrow P(\hat{x}|\omega_1) \cdot P(\omega_1) = P(\hat{x}|\omega_2) \cdot P(\omega_2)$$

$$\Rightarrow P(\hat{x}|\omega_1) = P(\hat{x}|\omega_2), \quad \text{for } P(\omega_1) = P(\omega_2) = \tfrac{1}{2}$$

Case 1: $\hat{x} < 0$

$$P(\hat{x}|\omega_1) = P(\hat{x}|\omega_2) = 0, \quad \forall \hat{x} \in (-\infty, 0).$$

Case 2: $\hat{x} = 0$

$$P(\hat{x}|\omega_1) = \frac{\hat{x}}{\sigma_1^2} \cdot e^{\frac{-\hat{x}^2}{2\sigma_1^2}} = 0$$

$$P(\hat{x}|\omega_2) = \frac{\hat{x}}{\sigma_2^2} \cdot e^{\frac{-\hat{x}^2}{2\sigma_2^2}} = 0$$

$$\Rightarrow P(\hat{x}|\omega_1) = P(\hat{x}|\omega_2)$$

Case 3: $\hat{x} > 0$:

$$\frac{\hat{x}}{\sigma_1^2} \cdot e^{\frac{-\hat{x}^2}{2\sigma_1^2}} = \frac{\hat{x}}{\sigma_2^2} \cdot e^{\frac{-\hat{x}^2}{2\sigma_2^2}}$$

$$\Rightarrow e^{\frac{\hat{x}^2}{2\sigma_2^2} - \frac{\hat{x}^2}{2\sigma_1^2}} = \frac{\sigma_1^2}{\sigma_2^2}$$

$$\Rightarrow \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right) \cdot 2\hat{x}^2 = \ln \frac{\sigma_1^2}{\sigma_2^2}$$

$$\Rightarrow 2\hat{x}^2 \cdot \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 \sigma_2^2} = 2(\ln\sigma_1 - \ln\sigma_2)$$

$$\Rightarrow 2\hat{x}^2 = \frac{(\sigma_1\sigma_2)^2 \cdot 2(\ln\sigma_1 - \ln\sigma_2)}{\sigma_1^2 - \sigma_2^2}$$

$$\Rightarrow \hat{x} = \sigma_1\sigma_2 \sqrt{\frac{\ln\sigma_1 - \ln\sigma_2}{(\sigma_1^2 - \sigma_2^2)}} \quad (\hat{x} > 0)$$

In general, the decision boundary is

$$DB = (-\infty, 0] \cup \left\{ \sigma_1\sigma_2 \sqrt{\frac{\ln\sigma_1 - \ln\sigma_2}{\sigma_1^2 - \sigma_2^2}} \right\}$$

2.7 In a three-class, two-dimensional problem the feature vectors in each class are normally distributed with covariance matrix

$$\Sigma = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.8 \end{bmatrix}$$

The mean vectors for each class are $[0.1, 0.1]^T$, $[2.1, 1.9]^T$, $[-1.5, 2.0]^T$. Assuming that the classes are equiprobable, (a) classify the feature vector $[1.6, 1.5]^T$ according to the Bayes minimum error probability classifier; (b) draw the curves of equal Mahalanobis distance from $[2.1, 1.9]^T$.

$$\mu_1 = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 2.1 \\ 1.9 \end{bmatrix} \quad \mu_3 = \begin{bmatrix} -1.5 \\ 2.0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.8 \end{bmatrix}$$

(a) $$\Sigma^{-1} = \frac{1}{1.2\times1.8-0.4^2}\begin{bmatrix} 1.8 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} \qquad |\Sigma| = 2 \qquad d = 2$$

$$= \frac{1}{2}\begin{bmatrix} 1.8 & -0.4 \\ -0.4 & 1.2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}$$

Posterior Prob. Dist.

$$P(x|w_i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(x-\mu_i)^T \Sigma^{-1}(x-\mu_i)}$$

$$= \frac{1}{2\pi\sqrt{2}} e^{-\frac{1}{2}(x-\mu_i)^T \begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}(x-\mu_i)}$$

$$= \frac{1}{2\pi\sqrt{2}} e^{-d_M^2}$$

Calculate Squared Mahalanobis distance from the test point to each mean.

$$d_{M_1}^2 = [1.6-0.1 \quad 1.5-0.1]\begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}\begin{bmatrix} 1.6-0.1 \\ 1.5-0.1 \end{bmatrix}$$

$$= [1.5 \quad 1.4]\begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}\begin{bmatrix} 1.5 \\ 1.4 \end{bmatrix}$$

$$= [1.07 \quad 0.54]\begin{bmatrix} 1.5 \\ 1.4 \end{bmatrix}$$

$$= 2.561$$

$$d_{M_2}^2 = [1.6-2.1 \quad 1.5-1.9]\begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}\begin{bmatrix} 1.6-2.1 \\ 1.5-1.9 \end{bmatrix}$$

$$= [-0.5 \quad -0.4]\begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}\begin{bmatrix} -0.5 \\ -0.4 \end{bmatrix}$$

$$= [-0.37 \quad -0.14]\begin{bmatrix} -0.5 \\ -0.4 \end{bmatrix}$$

$$= 0.745$$

$$d_{M_3}^2 = [1.6+1.5 \quad 1.5-2.0]\begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}\begin{bmatrix} 1.6+1.5 \\ 1.5-2.0 \end{bmatrix}$$

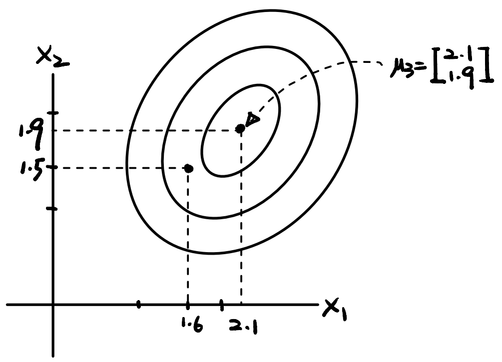
$$= [3.1 \quad -0.5]\begin{bmatrix} 0.9 & -0.2 \\ -0.2 & 0.6 \end{bmatrix}\begin{bmatrix} 3.1 \\ -0.5 \end{bmatrix}$$

$$= [2.89 \quad -0.92]\begin{bmatrix} 3.1 \\ -0.5 \end{bmatrix}$$

$$= 9.414$$

$d_{M_2}^2 < d_{M_1}^2 < d_{M_3}^2 \Rightarrow$ Classify to ω_2.

(b).



2.12 Consider a two-class, two-dimensional classification task, where the feature vectors in each of the classes ω_1, ω_2 are distributed according to

$$p(\boldsymbol{x}|\omega_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T(\boldsymbol{x}-\boldsymbol{\mu}_1)\right)$$

$$p(\boldsymbol{x}|\omega_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2\sigma_2^2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^T(\boldsymbol{x}-\boldsymbol{\mu}_2)\right)$$

with

$$\boldsymbol{\mu}_1 = [1, 1]^T, \quad \boldsymbol{\mu}_2 = [1.5, 1.5]^T, \quad \sigma_1^2 = \sigma_2^2 = 0.2$$

Assume that $P(\omega_1) = P(\omega_2)$ and design a Bayesian classifier

(a) that minimizes the error probability

(b) that minimizes the average risk with loss matrix

→ Ground Truth.

$$\Lambda = \begin{bmatrix} 0 & 1 \\ 0.5 & 0 \end{bmatrix} \lambda_{12}$$

Action ↓ λ_{21}

Using a pseudorandom number generator, produce 100 feature vectors from each class, according to the preceding pdfs. Use the classifiers designed to classify the generated vectors. What is the percentage error for each case? Repeat the experiments for $\boldsymbol{\mu}_2 = [3.0, 3.0]^T$.

(a) Minimize Error Prob.

$$P(\omega_1|x) = P(\omega_2|x)$$

$$\Rightarrow P(x|\omega_1)\cdot P(\omega_1) = P(x|\omega_2)\cdot P(\omega_2)$$

$$\Rightarrow P(x|\omega_1) = P(x|\omega_2)$$

$$\Rightarrow \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^T(x-\mu_1)} = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^T(x-\mu_2)}$$

$$\Rightarrow \frac{\sigma_2}{\sigma_1} e^{\frac{1}{2\sigma_2^2}(x-\mu_2)^T(x-\mu_2) - \frac{1}{2\sigma_1^2}(x-\mu_1)^T(x-\mu_1)} = 1$$

$$\Rightarrow \frac{1}{2\sigma_2^2}(x-\mu_2)^T(x-\mu_2) - \frac{1}{2\sigma_1^2}(x-\mu_1)^T(x-\mu_1) = \ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow (x-\mu_2)^T(x-\mu_2) - (x-\mu_1)^T(x-\mu_1) = 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow \left[(x-\mu_2)+(x-\mu_1)\right]^T\left[(x-\mu_2)-(x-\mu_1)\right] = 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow \left[2x-(\mu_1+\mu_2)\right]^T(\mu_1-\mu_2) = 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow \left[2\begin{pmatrix}x_1\\x_2\end{pmatrix} - \begin{pmatrix}2.5\\2.5\end{pmatrix}\right]^T \cdot \begin{pmatrix}-0.5\\-0.5\end{pmatrix} = 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow (2x_1-2.5 \quad 2x_2-2.5)\begin{pmatrix}-0.5\\-0.5\end{pmatrix} = 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow (-x_1+1.25)+(-x_2+1.25) = 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow 2.5-(x_1+x_2) = 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$\Rightarrow (x_1+x_2) = 2.5 - 0.4\ln\frac{\sigma_1}{\sigma_2}$$

$$= 2.5$$

(b) Minimize Average Risk.

$$R(\alpha_1|x) = R(\alpha_2|x)$$

$$\Rightarrow \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) = \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x)$$

$$\Rightarrow \lambda_{12}\cdot P(\omega_2|x) = \lambda_{21}\cdot P(\omega_1|x)$$

$$\Rightarrow \lambda_{12}\cdot P(x|\omega_2)\cdot P(\omega_2) = \lambda_{21}\cdot P(x|\omega_1)\cdot P(\omega_1)$$

$$\Rightarrow \lambda_{12}\cdot P(x|\omega_2) = \lambda_{21}\cdot P(x|\omega_1)$$

$$\Rightarrow \frac{P(x|\omega_1)}{P(x|\omega_2)} = \frac{\lambda_{12}}{\lambda_{21}} = \frac{1}{0.5} = 2$$

$$\Rightarrow \frac{\frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^T(x-\mu_1)}}{\frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^T(x-\mu_2)}} = 2$$

$$\Rightarrow \frac{1}{0.4}\left[(x-\mu_2)^T(x-\mu_2) - (x-\mu_1)^T(x-\mu_1)\right] = \ln 2$$

$$\Rightarrow \left[(x-\mu_2)+(x-\mu_1)\right]^T\left[(x-\mu_2)-(x-\mu_1)\right] = 0.4\ln 2$$

$$\Rightarrow \left[2x-(\mu_1+\mu_2)\right]^T(\mu_1-\mu_2) = 0.4\ln 2$$

$$\Rightarrow \left[2\begin{pmatrix}x_1\\x_2\end{pmatrix} - \begin{pmatrix}2.5\\2.5\end{pmatrix}\right]^T\begin{pmatrix}-0.5\\-0.5\end{pmatrix} = 0.4\ln 2$$

$$\Rightarrow \left[2x_1-2.5 \quad 2x_2-2.5\right]\begin{pmatrix}-0.5\\-0.5\end{pmatrix} = 0.4\ln 2$$

$$\Rightarrow (-x_1+1.25)+(-x_2+1.25) = 0.4\ln 2$$

$$\Rightarrow -(x_1+x_2)+2.5 = 0.4\ln 2$$

$$\Rightarrow x_1+x_2 = 2.5-0.4\ln 2$$

CISC3024 Final Review - CH3. Estimation Methods.

2.17 In a heads or tails coin-tossing experiment the probability of occurrence of a head (1) is q and that of a tail (0) is $1-q$. Let $x_i, i = 1, 2, \ldots, N$, be the resulting experimental outcomes, $x_i \in \{0, 1\}$. Show that the ML estimate of q is

$$q_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Hint: The likelihood function is

$$P(X : q) = \prod_{i=1}^{N} q^{x_i}(1-q)^{(1-x_i)}$$

Then show that the ML results from the solution of the equation

$$q^{\sum_i x_i}(1-q)^{(N-\sum_i x_i)}\left(\frac{\sum_i x_i}{q} - \frac{N-\sum_i x_i}{1-q}\right) = 0$$

A single likelihood:

$$P(x_i|q) = \begin{cases} q & x_i = 1 \\ 1-q & x_i = 0 \end{cases} = q^{x_i}(1-q)^{1-x_i}$$

Total likelihood:

$$L(q) = \prod_{i=1}^{N} P(x_i|q)$$

$$= \prod_{i=1}^{N} q^{x_i}(1-q)^{1-x_i}$$

Log likelihood:

$$\mathcal{L}(q) = \ln L(q)$$

$$= \ln \prod_{i=1}^{N} q^{x_i}(1-q)^{1-x_i}$$

$$= \sum_{i=1}^{N} \ln(q^{x_i}(1-q)^{1-x_i})$$

$$= \sum_{i=1}^{N} x_i \cdot \ln q + (1-x_i)\ln(1-q).$$

Minimize log likelihood:

$$\frac{\partial \mathcal{L}}{\partial q} = 0 \Rightarrow \sum_{i=1}^{N} \frac{x_i}{q} - \frac{1-x_i}{1-q} = 0$$

$$\Rightarrow \sum_{i=1}^{N} (1-q)x_i - q(1-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{N} x_i - qx_i - q + qx_i = 0$$

$$\Rightarrow \sum_{i=1}^{N} x_i - q = 0$$

$$\Rightarrow \left(\sum_{i=1}^{N} x_i\right) - N \cdot q = 0$$

$$\Rightarrow q = \frac{1}{N}\sum_{i=1}^{N} x_i$$

2.29 Show that for the lognormal distribution

$$p(x) = \frac{1}{\sigma x\sqrt{2\pi}}\exp\left(-\frac{(\ln x - \theta)^2}{2\sigma^2}\right), \quad x > 0$$

the ML estimate is given by

$$\hat{\theta}_{ML} = \frac{1}{N}\sum_{k=1}^{N} \ln x_k$$

Single Likelihood:

$$P(x|\theta) = \frac{1}{\sigma\sqrt{2\pi}\cdot x} e^{-\frac{(\ln x - \theta)^2}{2\sigma^2}}$$

Logged Single Likelihood:

$$h(\theta) = \ln P(x|\theta)$$

$$= -\ln(\sigma\sqrt{2\pi}x) - \frac{(\ln x - \theta)^2}{2\sigma^2}$$

$$\frac{\partial h}{\partial \theta} = \frac{\partial}{\partial \theta} - \frac{(\ln x - \theta)^2}{2\sigma^2}$$

$$= -\frac{1}{2\sigma^2}\cdot\frac{\partial}{\partial\theta}\cdot(\theta^2 - 2\theta\ln x + \ln^2 x)$$

$$= -\frac{1}{2\sigma^2}\cdot(2\theta - 2\ln x)$$

$$= \frac{\ln x - \theta}{\sigma^2}$$

Likelihood:

$$L(\theta) = \prod_{k=1}^{N} P(x|\theta)$$

$$= \prod_{k=1}^{N} \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x - \theta)^2}{2\sigma^2}}$$

Logged likelihood:

$$\mathcal{L}(\theta) = \ln \prod_{k=1}^{N} P(x|\theta)$$

$$= \sum_{k=1}^{N} \ln P(x|\theta)$$

$$= \sum_{k=1}^{N} h(\theta)$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0$$

$$\Rightarrow \frac{\partial}{\partial\theta}\sum_{k=1}^{N} h(\theta) = 0$$

$$\Rightarrow \sum_{k=1}^{N} \frac{\partial}{\partial\theta} h(\theta) = 0$$

$$\Rightarrow \sum_{k=1}^{N} \frac{1}{\sigma^2}(\ln x_k - \theta) = 0$$

$$\Rightarrow \frac{1}{\sigma^2}\sum_{k=1}^{N}(\ln x_k - \theta) = 0$$

$$\Rightarrow \sum_{k=1}^{N} \ln x_k - \theta = 0$$

$$\Rightarrow \left(\sum_{k=1}^{N} \ln x_k\right) - N\theta = 0$$

$$\Rightarrow \theta = \frac{1}{N}\sum_{k=1}^{N} \ln x_k.$$

Q1. Perform SVD on $A = \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix}$

Step 1. Find AA^T and A^TA.

$A^T = \begin{pmatrix} 4 & 3 \\ 0 & -5 \end{pmatrix}$

$\Rightarrow AA^T = \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix}\begin{pmatrix} 4 & 3 \\ 0 & -5 \end{pmatrix} = \begin{pmatrix} 16 & 12 \\ 12 & 34 \end{pmatrix}$

$A^TA = \begin{pmatrix} 4 & 3 \\ 0 & -5 \end{pmatrix}\begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix} = \begin{pmatrix} 25 & -15 \\ -15 & 25 \end{pmatrix}$

Step 2. Eigenvalue and S

$AA^T = \lambda I$

$\Rightarrow |AA^T - \lambda I| = 0$

$\Rightarrow \left|\begin{pmatrix} 16 & 12 \\ 12 & 34 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right| = 0$

$\Rightarrow \begin{vmatrix} 16-\lambda & 12 \\ 12 & 34-\lambda \end{vmatrix} = 0$

$\Rightarrow (\lambda-16)(\lambda-34) - 12^2 = 0$

$\Rightarrow \lambda^2 - 50\lambda + 16\times 34 - 12^2 = 0$

$\Rightarrow \lambda^2 - 50\lambda + 400 = 0$

$\Rightarrow (\lambda - 40)(\lambda - 10) = 0$

$\Rightarrow \lambda_1 = 40, \ \lambda_2 = 10$

$\Rightarrow \sigma_1 = \sqrt{40} = 2\sqrt{10}, \ \sigma_2 = \sqrt{10}$

$\therefore S = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = \begin{pmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{pmatrix}$

Step 3. Find U.

$(AA^T - \lambda I)x = 0$

① For $\lambda_1 = 40$

$\Rightarrow \left[\begin{pmatrix} 16 & 12 \\ 12 & 34 \end{pmatrix} - \begin{pmatrix} 40 & 0 \\ 0 & 40 \end{pmatrix}\right] x_1 = 0$

$\Rightarrow \begin{pmatrix} -24 & 12 \\ 12 & -6 \end{pmatrix} x_1 = 0$

$\Rightarrow \left[\begin{array}{cc|c} -24 & 12 & 0 \\ 12 & -6 & 0 \end{array}\right]$

$\xrightarrow{r_2 + \frac{1}{2}r_1} \left[\begin{array}{cc|c} -24 & 12 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\xrightarrow{r_1 \cdot -\frac{1}{12}} \left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

...

$\Rightarrow 2x_{11} - x_{12} = 0$

$\Rightarrow x_{12} = 2x_{11}$

$\Rightarrow x_1 = a_1\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ a_1 \in \mathbb{R}, a_1 \neq 0$

$u_1 = \frac{x_1}{\|x_1\|} = \frac{1}{\sqrt{1+2^2}} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$

② For $\lambda_2 = 10$,

$\left[\begin{pmatrix} 16 & 12 \\ 12 & 34 \end{pmatrix} - \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}\right] x_2 = 0$

$\Rightarrow \begin{pmatrix} 6 & 12 \\ 12 & 24 \end{pmatrix} x_2 = 0$

$\Rightarrow \left[\begin{array}{cc|c} 6 & 12 & 0 \\ 12 & 24 & 0 \end{array}\right]$

$\xrightarrow{r_2 - 2r_1} \left[\begin{array}{cc|c} 6 & 12 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\xrightarrow{r_1 \cdot \frac{1}{6}} \left[\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\Rightarrow x_{21} + 2x_{22} = 0$

$\Rightarrow x_{21} = -2x_{22}$

$\Rightarrow x_2 = a_2\begin{bmatrix} -2 \\ 1 \end{bmatrix}$

$u_2 = \frac{x_2}{\|x_2\|} = \frac{1}{\sqrt{2^2+1^2}}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$

In summary:

$U = \begin{pmatrix} | & | \\ u_1 & u_2 \\ | & | \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$

Step 4. Find V.

$(A^TA - \lambda I)x = 0$

① For $\lambda_1 = 40$

$\left[\begin{pmatrix} 25 & -15 \\ -15 & 25 \end{pmatrix} - \begin{pmatrix} 40 & 0 \\ 0 & 40 \end{pmatrix}\right] x_3 = 0$

$\Rightarrow \begin{pmatrix} -15 & -15 \\ -15 & -15 \end{pmatrix} x_3 = 0$

$\Rightarrow \left[\begin{array}{cc|c} -15 & -15 & 0 \\ -15 & -15 & 0 \end{array}\right]$

$\Rightarrow \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \end{array}\right]$

...

$\Rightarrow \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\Rightarrow x_{31} + x_{32} = 0$

$\Rightarrow x_{32} = -x_{31}$

$\Rightarrow x_3 = a_3\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \ a_3 \in \mathbb{R}, \ a_3 \neq 0,$

$v_1 = \frac{x_3}{\|x_3\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$

② For $\lambda_2 = 10$.

$\left[\begin{pmatrix} 25 & -15 \\ -15 & 25 \end{pmatrix} - \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}\right] x_4 = 0$

$\Rightarrow \begin{pmatrix} 15 & -15 \\ -15 & 15 \end{pmatrix} x_4 = 0$

$\Rightarrow \left[\begin{array}{cc|c} 15 & -15 & 0 \\ -15 & 15 & 0 \end{array}\right]$

$\rightarrow \left[\begin{array}{cc|c} 1 & -1 & 0 \\ -1 & 1 & 0 \end{array}\right]$

$\rightarrow \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

$\Rightarrow x_{41} - x_{42} = 0$

$\Rightarrow x_{41} = x_{42}$

$\Rightarrow x_4 = a_4\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ a_4 \in \mathbb{R}, \ a_4 \neq 0.$

$v_2 = \frac{x_4}{\|x_4\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

$\Rightarrow V = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$

Conclusion

$$\underbrace{\begin{pmatrix} 4 & 3 \\ 0 & -5 \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{pmatrix}}_{S} \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{V^T}$$

is the Singular Value Decomposition of A.

Q2. Perform SVD on $B=\begin{pmatrix}2&0&1\\-1&2&0\end{pmatrix}$

Step 1. Find BB^T and B^TB

$BB^T=\begin{pmatrix}2&0&1\\-1&2&0\end{pmatrix}\begin{pmatrix}2&-1\\0&2\\1&0\end{pmatrix}=\begin{pmatrix}5&-2\\-2&5\end{pmatrix}$

$B^TB=\begin{pmatrix}2&-1\\0&2\\1&0\end{pmatrix}\begin{pmatrix}2&0&1\\-1&2&0\end{pmatrix}=\begin{pmatrix}5&-2&2\\-2&4&0\\2&0&1\end{pmatrix}$

Step 2. Eigen Values and S.

$|BB^T-\lambda I|=0$

$\Rightarrow \left|\begin{pmatrix}5&-2\\-2&5\end{pmatrix}-\begin{pmatrix}\lambda&0\\0&\lambda\end{pmatrix}\right|=0$

$\Rightarrow \begin{vmatrix}5-\lambda&-2\\-2&5-\lambda\end{vmatrix}=0$

$\Rightarrow (5-\lambda)^2-4=0$

$\Rightarrow \lambda^2-10\lambda+25-4=0$

$\Rightarrow \lambda^2-10\lambda+21=0$

$\Rightarrow (\lambda-3)(\lambda-7)=0$

$\Rightarrow \lambda_1=7,\ \lambda_2=3,\ \lambda_3=0.$

$\Rightarrow \sigma_1=\sqrt{7},\ \sigma_2=\sqrt{3}$

$S=\begin{pmatrix}\sqrt{7}&0&0\\0&\sqrt{3}&0\end{pmatrix}$.

Step 3. Find U.

$(BB^T-\lambda I)x=0.$

① $\lambda_1=7.$

$\left[\begin{pmatrix}5&-2\\-2&5\end{pmatrix}-\begin{pmatrix}7&0\\0&7\end{pmatrix}\right]x=0$

$\Rightarrow \begin{pmatrix}-2&-2\\-2&-2\end{pmatrix}x_1=0$

$\Rightarrow \left[\begin{array}{cc|c}-2&-2&0\\-2&-2&0\end{array}\right]$

$\rightarrow \left[\begin{array}{cc|c}1&1&0\\1&1&0\end{array}\right]$

$\rightarrow \left[\begin{array}{cc|c}1&1&0\\0&0&0\end{array}\right]$

$\Rightarrow x_{11}+x_{12}=0$

$\Rightarrow x_{11}=-x_{12}$

$\Rightarrow x_1=a_1\begin{bmatrix}-1\\1\end{bmatrix}, a_1\in\mathbb{R}, a_1\neq 0.$

$\Rightarrow u_1=\frac{x_1}{\|x_1\|}=\begin{pmatrix}-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}.$

② $\lambda_2=3$

$\left[\begin{pmatrix}5&-2\\-2&5\end{pmatrix}-\begin{pmatrix}3&0\\0&3\end{pmatrix}\right]x_2=0$

$\Rightarrow \begin{pmatrix}2&-2\\-2&2\end{pmatrix}x_2=0$

$\Rightarrow \left[\begin{array}{cc|c}2&-2&0\\-2&2&0\end{array}\right]$

$\rightarrow \left[\begin{array}{cc|c}1&-1&0\\-1&1&0\end{array}\right]$

$\rightarrow \left[\begin{array}{cc|c}1&-1&0\\0&0&0\end{array}\right]$

$\Rightarrow x_{21}-x_{22}=0$

$\Rightarrow x_{21}=x_{22}$

$\Rightarrow x_2=a_2\begin{bmatrix}1\\1\end{bmatrix}, a_2\in\mathbb{R}. \ a_2\neq 0.$

$\Rightarrow u_2=\frac{x_2}{\|x_2\|}=\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}$

$\therefore U=(u_1\ u_2)=\begin{pmatrix}-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix}$

Step 4. Find V.

$(B^TB-\lambda I)x=0.$

① $\lambda_1=7$

$\left[\begin{pmatrix}5&-2&2\\-2&4&0\\2&0&1\end{pmatrix}-\begin{pmatrix}7&0&0\\0&7&0\\0&0&7\end{pmatrix}\right]x=0.$

$\Rightarrow \begin{pmatrix}-2&-2&2\\-2&-3&0\\2&0&-6\end{pmatrix}x=0$

$\Rightarrow \left[\begin{array}{ccc|c}-2&-2&2&0\\-2&-3&0&0\\2&0&-6&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}2&2&-2&0\\2&3&0&0\\2&0&-6&0\end{array}\right]$

$\xrightarrow[r_3-r_1]{r_2-r_1} \left[\begin{array}{ccc|c}2&2&-2&0\\0&1&2&0\\0&-2&-4&0\end{array}\right]$

$\xrightarrow[r_3\cdot(-1)]{r_2\cdot 2} \left[\begin{array}{ccc|c}2&2&-2&0\\0&2&4&0\\0&2&4&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}2&2&-2&0\\0&2&4&0\\0&0&0&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}1&1&-1&0\\0&1&2&0\\0&0&0&0\end{array}\right]$

$\xrightarrow{r_1-r_2} \left[\begin{array}{ccc|c}1&0&-3&0\\0&1&2&0\\0&0&0&0\end{array}\right]$

$\begin{cases}x_{31}-3x_{33}=0.\\x_{32}+2x_{33}=0\end{cases}$

$\Rightarrow \begin{cases}x_{31}=3x_{33}\\x_{32}=-2x_{33}\end{cases}$

$\Rightarrow x_3=a_3\begin{bmatrix}3\\-2\\1\end{bmatrix}, a_3\in\mathbb{R}, a_3\neq 0.$

$\Rightarrow v_1=\frac{x_3}{\|x_3\|}=\begin{pmatrix}\frac{3}{\sqrt{14}}\\-\frac{2}{\sqrt{14}}\\\frac{1}{\sqrt{14}}\end{pmatrix}$

② $\lambda_2=3$

$\left[\begin{pmatrix}5&-2&2\\-2&4&0\\2&0&1\end{pmatrix}-\begin{pmatrix}3&0&0\\0&3&0\\0&0&3\end{pmatrix}\right]x_4=0$

$\Rightarrow \begin{pmatrix}2&-2&2\\-2&1&0\\2&0&-2\end{pmatrix}x_4=0$

$\Rightarrow \left[\begin{array}{ccc|c}2&-2&2&0\\-2&1&0&0\\2&0&-2&0\end{array}\right]$

$\Rightarrow \left[\begin{array}{ccc|c}2&-2&2&0\\2&-1&0&0\\2&0&-2&0\end{array}\right]$

$\xrightarrow[r_3-r_1]{r_2-r_1} \left[\begin{array}{ccc|c}2&-2&2&0\\0&1&-2&0\\0&2&-4&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}2&-2&2&0\\0&1&-2&0\\0&0&0&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}1&-1&1&0\\0&1&-2&0\\0&0&0&0\end{array}\right]$

$\xrightarrow{r_1+r_2} \left[\begin{array}{ccc|c}1&0&-1&0\\0&1&-2&0\\0&0&0&0\end{array}\right]$

$\Rightarrow \begin{cases}x_{41}-x_{43}=0\\x_{42}-2x_{43}=0\end{cases}$

$\Rightarrow \begin{cases}x_{41}=x_{43}\\x_{42}=2x_{43}\end{cases}$

$\Rightarrow x_4=a_4\begin{bmatrix}1\\2\\1\end{bmatrix}, a_4\in\mathbb{R}, a_4\neq 0$

$\Rightarrow v_2=\frac{x_4}{\|x_4\|}=\begin{pmatrix}\frac{1}{\sqrt{6}}\\\frac{2}{\sqrt{6}}\\\frac{1}{\sqrt{6}}\end{pmatrix}$

③ $\lambda_3=0.$

$\left[\begin{pmatrix}5&-2&2\\-2&4&0\\2&0&1\end{pmatrix}-\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}\right]x=0$

$\Rightarrow \begin{pmatrix}5&-2&2\\-2&4&0\\2&0&1\end{pmatrix}x=0$

$\Rightarrow \left[\begin{array}{ccc|c}5&-2&2&0\\-2&4&0&0\\2&0&1&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}10&-4&4&0\\10&-20&0&0\\10&0&5&0\end{array}\right]$

$\xrightarrow[r_3-r_1]{r_2-r_1} \left[\begin{array}{ccc|c}10&-4&4&0\\0&-16&-4&0\\0&4&1&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}10&-4&4&0\\0&4&1&0\\0&0&0&0\end{array}\right]$

$\xrightarrow{r_1+r_2} \left[\begin{array}{ccc|c}10&0&5&0\\0&4&1&0\\0&0&0&0\end{array}\right]$

$\rightarrow \left[\begin{array}{ccc|c}2&0&1&0\\0&4&1&0\\0&0&0&0\end{array}\right]$

$\Rightarrow \begin{cases}2x_{41}+x_{43}=0\\4x_{42}+x_{43}=0\end{cases}$

$\Rightarrow \begin{cases}x_{41}=2x_{42}\\x_{43}=-4x_{42}\end{cases}$

$\Rightarrow x_4=a_4\begin{pmatrix}2\\1\\-4\end{pmatrix}\ a_4\in\mathbb{R}\ \ a_4\neq 0$

$\Rightarrow v_3=\frac{x_4}{\|x_4\|}=\begin{pmatrix}\frac{2}{\sqrt{21}}\\\frac{1}{\sqrt{21}}\\-\frac{4}{\sqrt{21}}\end{pmatrix}$

$\therefore V=(v_1\ v_2\ v_3)$

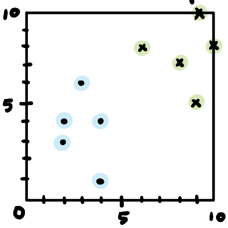
$=\begin{pmatrix}\frac{3}{\sqrt{14}}&\frac{1}{\sqrt{6}}&\frac{2}{\sqrt{21}}\\-\frac{2}{\sqrt{14}}&\frac{2}{\sqrt{6}}&\frac{1}{\sqrt{21}}\\\frac{1}{\sqrt{14}}&\frac{1}{\sqrt{6}}&-\frac{4}{\sqrt{21}}\end{pmatrix}$

Conclusion

$$\underset{B}{\begin{pmatrix}2&0&1\\-1&2&0\end{pmatrix}}$$

$$=\underset{U}{\begin{pmatrix}-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{pmatrix}}\underset{S}{\begin{pmatrix}\sqrt{7}&0&0\\0&\sqrt{3}&0\end{pmatrix}}\underset{V^T}{\begin{pmatrix}\frac{3}{\sqrt{14}}&-\frac{2}{\sqrt{14}}&\frac{1}{\sqrt{14}}\\\frac{1}{\sqrt{6}}&\frac{2}{\sqrt{6}}&\frac{1}{\sqrt{6}}\\\frac{2}{\sqrt{21}}&\frac{1}{\sqrt{21}}&-\frac{4}{\sqrt{21}}\end{pmatrix}}$$

Q3. Perform LDA on the following sample:



$X_1 = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$;

$X_2 = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$.

Step 1. Data Arrange

$$\chi_1 = \begin{pmatrix} 4 & 2 & 2 & 3 & 4 \\ 1 & 4 & 3 & 6 & 4 \end{pmatrix}$$

$$\chi_2 = \begin{pmatrix} 9 & 6 & 9 & 8 & 10 \\ 10 & 8 & 5 & 7 & 8 \end{pmatrix}$$

Step 2. Cls Statistics.

① Sample Means.

$$\mu_1 = \begin{pmatrix} \frac{1}{5}(4+2+2+3+4) \\ \frac{1}{5}(1+4+3+6+4) \end{pmatrix} = \begin{pmatrix} 3 \\ 3.6 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} \frac{1}{5}(9+6+9+8+10) \\ \frac{1}{5}(10+8+5+7+8) \end{pmatrix} = \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix}$$

② Sample Variants.

$$S_1 = \frac{1}{5}\begin{pmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{pmatrix}\begin{pmatrix} 1 & -2.6 \\ -1 & 0.4 \\ -1 & -0.6 \\ 0 & 2.4 \\ 1 & 0.4 \end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix} 4 & -2 \\ -2 & 13.2 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{pmatrix}$$

$$S_2 = \frac{1}{5}\begin{pmatrix} 0.6 & -2.4 & 0.6 & -0.4 & 1.6 \\ 2.4 & 0.4 & -2.6 & -0.6 & 0.4 \end{pmatrix}\begin{pmatrix} 0.6 & 2.4 \\ -2.4 & 0.4 \\ 0.6 & -2.6 \\ -0.4 & -0.6 \\ 1.6 & 0.4 \end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix} 9.2 & -0.2 \\ -0.2 & 13.2 \end{pmatrix}$$

$$= \begin{pmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{pmatrix}$$

Step 3. Between-Class Scatters.

$$S_W = S_1 + S_2$$

$$= \begin{pmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{pmatrix} + \begin{pmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{pmatrix}$$

$$= \begin{pmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{pmatrix}$$

Step 4. Optimum Weight.

$$S_W^{-1} = \frac{1}{2.64 \times 5.28 - 0.44^2}\begin{pmatrix} 5.28 & 0.44 \\ 0.44 & 2.64 \end{pmatrix}$$

$$= \frac{1}{13.7456}\begin{pmatrix} 5.28 & 0.44 \\ 0.44 & 2.64 \end{pmatrix}$$

$$= \begin{pmatrix} 0.3841 & 0.0320 \\ 0.0320 & 0.1921 \end{pmatrix}$$

$$\mu_1 - \mu_2 = \begin{pmatrix} 3 \\ 3.6 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix} = \begin{pmatrix} -5.4 \\ -4 \end{pmatrix}$$

$$w^* = S_W^{-1}(\mu_1 - \mu_2)$$

$$= \begin{pmatrix} 0.3841 & 0.0320 \\ 0.0320 & 0.1921 \end{pmatrix}\begin{pmatrix} -5.4 \\ -4 \end{pmatrix}$$

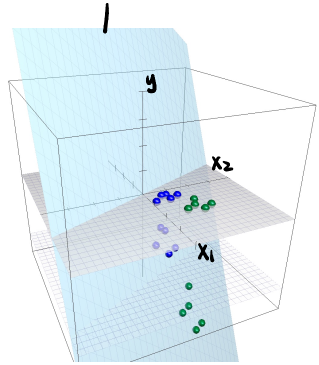
$$= \begin{pmatrix} -2.2021 \\ -0.9412 \end{pmatrix}$$

Conclusion

The line would be:

$$y = \begin{pmatrix} -2.2021 & -0.9412 \end{pmatrix} x$$

$$y = \begin{pmatrix} -2.2021 & -0.9412 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



CISC3024 Final Review - CH05. SVM

Q1. We have 5 1-d data points.

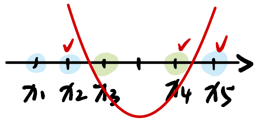
$x_1 = 1 \in \omega_1$

$x_2 = 2 \in \omega_1$

$x_3 = 3 \in \omega_2$

$x_4 = 5 \in \omega_2$

$x_5 = 6 \in \omega_1$



- Set degree-2 polynomial kernel: $K(x_i, x_j) = (x_i \cdot x_j + 1)^2$
- Set parameter $C = 100$.

Find Lagrangian polynomial $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ s.t.:

$$G(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\lambda_i \cdot \lambda_j) \cdot (y_i \cdot y_j) \cdot K(x_i, x_j)$$

is maximized.

$$\Rightarrow G(\lambda) = \sum_{i=1}^{5} \lambda_i - \frac{1}{2} \sum_{i=1}^{5} \sum_{j=1}^{5} (\lambda_i \cdot \lambda_j) \cdot (y_i \cdot y_j) \cdot (x_i x_j + 1)^2$$

with constraint:

$$\begin{cases} \sum_{i=1}^{5} \lambda_i y_i = 0 \\ 0 \le \lambda_i \le 100. \end{cases}$$

From quadratic problem solver:

$\lambda_1 = 0,\ \lambda_2 = 2.5,\ \lambda_3 = 0,\ \lambda_4 = 7.333,\ \lambda_5 = 4.833$

Support vectors:
$$\begin{cases} x_2 = 2 \\ x_4 = 5 \\ x_5 = 6 \end{cases}$$

Solution:

$$f(x) = \sum_{i=1}^{N} \lambda_i y_i K(x_i, x) + b$$

$$= \lambda_2 y_2 K(x_2, x) + \lambda_4 y_4 K(x_5, x) + \lambda_5 y_5 K(x_6, x) + b.$$

$$= 2.5 \times 1 \times (2x+1)^2 + 7.333 \times (-1) \times (5x+1)^2 + 4.833 \times 1 \times (6x+1)^2 + b.$$

$$= 2.5 \times (4x^2 + 4x + 1) - 7.333(25x^2 + 10x + 1) + 4.833(36x^2 + 12x + 1) + b.$$

$$= 10x^2 + 10x + 2.5 - 183.325x^2 - 73.33x - 7.333 + 173.988x^2 + 57.996x + 4.833 + b$$

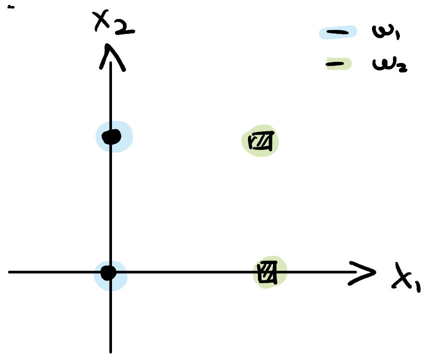
$$= 0.6667y^2 - 5.333y + b$$

$f(6) = 1 \Rightarrow b = 9$

Conclusion:

$$f(x) = 0.6667y^2 - 5.333y + 9.$$

3.4 Consider a case in which class ω_1 consists of the two feature vectors $[0,0]^T$ and $[0,1]^T$ and class ω_2 of $[1,0]^T$ and $[1,1]^T$. Use the perceptron algorithm in its reward and punishment form, with $\rho = 1$ and $w(0) = [0,0]^T$, to design the line separating the two classes.



Perceptron Algorithm:

$$w_{t+1} = w_t - \eta \cdot \sum_{x \in Y} \delta_x \cdot x$$

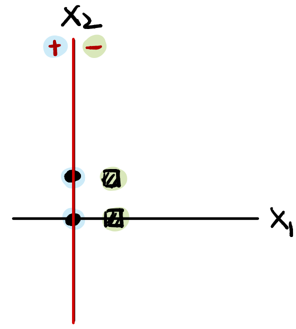
$$y_{pred} = \begin{cases} w_1 & w^T x > 0 \\ \text{unknown} & w^T x = 0 \\ w_2 & w^T x < 0 \end{cases}$$

Let $\delta_x = \begin{cases} -1 & x \in \omega_1 \\ 1 & x \in \omega_2 \end{cases}$

Iteration (t)	W	x	$w^T x$	y_{pred}	y_{truth}	$x \in Y$?
0	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	0	?	w_1	×
		$\begin{bmatrix}0\\1\\1\end{bmatrix}$	0	?	w_1	×
		$\begin{bmatrix}1\\0\\1\end{bmatrix}$	0	?	w_2	×
		$\begin{bmatrix}1\\1\\1\end{bmatrix}$	0	?	w_2	×

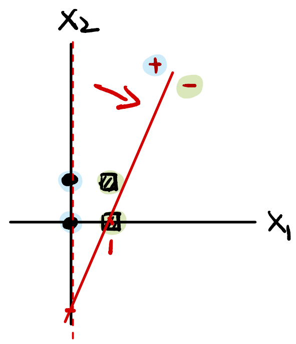
Update:

$$w_1 = \begin{bmatrix}0\\0\\0\end{bmatrix} - 1\times\left(-\begin{bmatrix}0\\0\\1\end{bmatrix} - \begin{bmatrix}0\\1\\1\end{bmatrix} + \begin{bmatrix}1\\0\\1\end{bmatrix} + \begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}0\\0\\0\end{bmatrix} - \begin{bmatrix}-0-0+1+1\\-0-1+0+1\\-1-1+1+1\end{bmatrix} = \begin{bmatrix}-2\\0\\0\end{bmatrix}$$



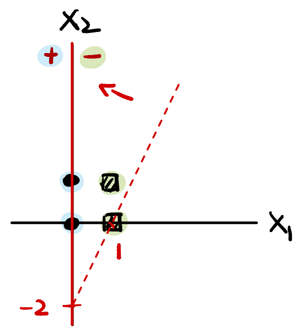
Iteration (t)	W	x	$w^T x$	y_{pred}	y_{truth}	$x \in Y$?
1	$\begin{bmatrix}-2\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	0	?	w_1	×
		$\begin{bmatrix}0\\1\\1\end{bmatrix}$	0	?	w_1	×
		$\begin{bmatrix}1\\0\\1\end{bmatrix}$	−2	w_2	w_2	✓
		$\begin{bmatrix}1\\1\\1\end{bmatrix}$	−2	w_2	w_2	✓

$$w_2 = \begin{bmatrix}-2\\0\\0\end{bmatrix} - 1\times\left(-\begin{bmatrix}0\\0\\1\end{bmatrix} - \begin{bmatrix}0\\1\\1\end{bmatrix}\right) = \begin{bmatrix}-2\\0\\0\end{bmatrix} - \begin{bmatrix}-0-0\\-0-1\\-1-1\end{bmatrix} = \begin{bmatrix}-2\\1\\2\end{bmatrix}$$



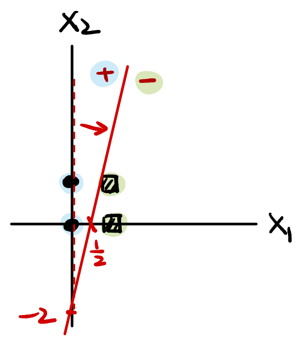
Iteration (t)	W	x	$w^T x$	y_{pred}	y_{truth}	$x \in Y$?
2	$\begin{bmatrix}-2\\1\\2\end{bmatrix}$	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	2	w_1	w_1	✓
		$\begin{bmatrix}0\\1\\1\end{bmatrix}$	3	w_1	w_1	✓
		$\begin{bmatrix}1\\0\\1\end{bmatrix}$	0	w_1	w_2	×
		$\begin{bmatrix}1\\1\\1\end{bmatrix}$	1	w_1	w_2	×

$$w_3 = \begin{bmatrix}-2\\1\\2\end{bmatrix} - 1\times\left(\begin{bmatrix}1\\0\\1\end{bmatrix} + \begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}-2\\1\\2\end{bmatrix} - \begin{bmatrix}1+1\\0+1\\1+1\end{bmatrix} = \begin{bmatrix}-4\\0\\0\end{bmatrix}$$



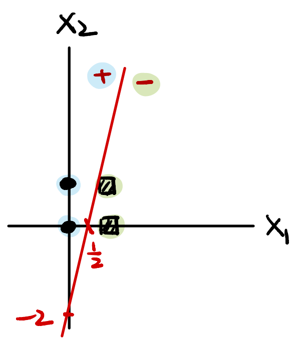
Iteration (t)	W	x	$w^T x$	y_{pred}	y_{truth}	$x \in Y$?
3	$\begin{bmatrix}-4\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	0	?	w_1	×
		$\begin{bmatrix}0\\1\\1\end{bmatrix}$	0	?	w_1	×
		$\begin{bmatrix}1\\0\\1\end{bmatrix}$	−4	w_2	w_2	✓
		$\begin{bmatrix}1\\1\\1\end{bmatrix}$	−4	w_2	w_2	✓

$$w_4 = \begin{bmatrix}-4\\0\\0\end{bmatrix} - 1\times\left(-\begin{bmatrix}0\\0\\1\end{bmatrix} - \begin{bmatrix}0\\1\\1\end{bmatrix}\right) = \begin{bmatrix}-4\\0\\0\end{bmatrix} - \begin{bmatrix}-0-0\\-0-1\\-1-1\end{bmatrix} = \begin{bmatrix}-4\\1\\2\end{bmatrix}$$



Iteration (t)	W	x	$w^T x$	y_{pred}	y_{truth}	$x \in Y$?
4	$\begin{bmatrix}-4\\1\\2\end{bmatrix}$	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	2	w_1	w_1	✓
		$\begin{bmatrix}0\\1\\1\end{bmatrix}$	3	w_1	w_1	✓
		$\begin{bmatrix}1\\0\\1\end{bmatrix}$	−2	w_2	w_2	✓
		$\begin{bmatrix}1\\1\\1\end{bmatrix}$	−1	w_2	w_2	✓

W converges.



4.3 Draw the three lines in the two-dimensional space

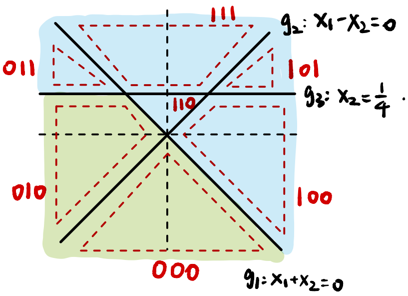
$$x_1 + x_2 = 0$$

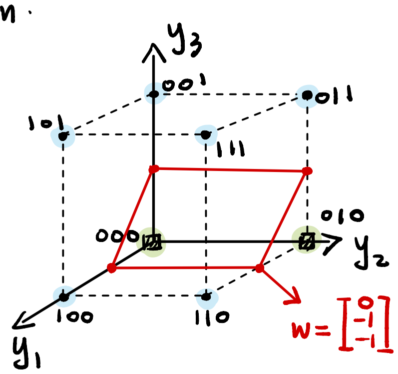
$$x_2 = \frac{1}{4}$$

$$x_1 - x_2 = 0$$

For each of the polyhedra that are formed by their intersections, determine the vertices of the cube into which they will be mapped by the first layer of a multilayer perceptron, realizing the preceding lines. Combine the regions into two classes so that (a) a two-layer network is sufficient to classify them and (b) a three-layer network is necessary. For both cases compute analytically the corresponding synaptic weights.

(a) Two-layer Perceptron.

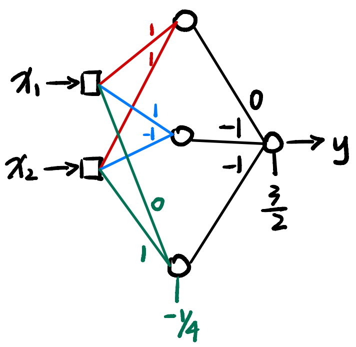




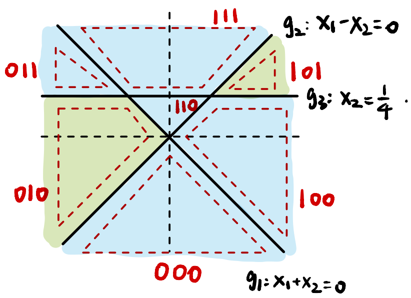
$W = \begin{bmatrix}0\\-1\\-1\end{bmatrix}$

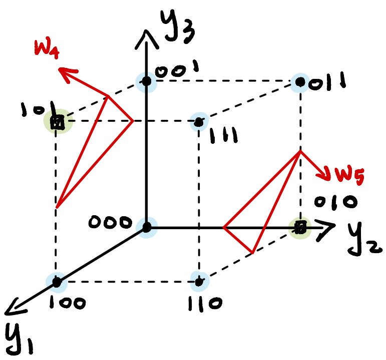
$[0\ \ -1\ \ -1\ \ b]\begin{bmatrix}\frac{1}{2}\\1\\0\\1\end{bmatrix} = 0$

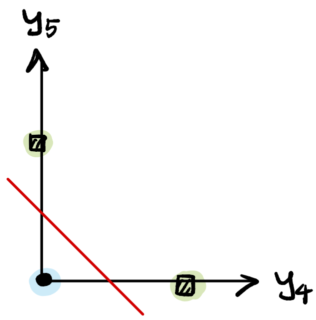
$\Rightarrow b = \frac{3}{2}$



(b) Three-layer Perceptron.







$W_6 = \begin{bmatrix}1\\1\end{bmatrix}$ $b_6 = -\frac{1}{2}$

$W_4 = \begin{bmatrix}1\\-1\\1\end{bmatrix}$, $[1\ \ -1\ \ 1\ \ b_4]\begin{bmatrix}1\\0\\\frac{1}{2}\\1\end{bmatrix} = 0 \Rightarrow b_4 = -\frac{3}{2}$

$W_5 = \begin{bmatrix}-1\\1\\-1\end{bmatrix}$, $[-1\ \ 1\ \ -1\ \ b_5]\begin{bmatrix}0\\1\\\frac{1}{2}\\1\end{bmatrix} = 0 \Rightarrow b_5 = -\frac{1}{2}$