Machine Learning I

Homework 05

Theory part , Fischer Discriminant

my solution

17. Nov. 2025

Machine Learning I   Exercise Sheet 5

Task 1

a) As given in the task description, we can extend the objective with a constraint.

this can be proven by : $\alpha \in \mathbb{R}$

$$f(w) = \frac{w^T S_B w}{w^T S_w w} \quad , \quad f(\alpha w) = \frac{(\alpha w)^T S_B (\alpha w)}{(\alpha w)^T S_w (\alpha w)} = \frac{\alpha^2 w^T S_B w}{\alpha^2 w^T S_w w} = \frac{w^T S_B w}{w^T S_w w}$$

therefore, we can constraint $w^T S_w w$ as a ~~cons~~ quadratic constraint.

Definition: $w^T S_w w = 1$

then the optimization problem can be reformulate as:

$$\max_w \frac{w^T S_B w}{w^T S_w w} = \max_w \frac{w^T S_B w}{1} = \max w^T S_B w$$

the quadratic objective : $w^T S_B w$

the quadratic constraint : $w^T S_w w = 1$.

(b) Lagrangian formula : $\mathcal{L}(w,\lambda) = f(w) - \lambda g(w)$

in task ~~1(a)~~ 1.(a) I have reformulate the problem. using the Reformulation above.

$w^T S_w w = 1$

$g(w) = w^T S_w w - 1$

$$L(w,\lambda) = f(w) - \lambda \cdot (w^T S_w w - 1)$$
$$= w^T S_B w - \lambda (w^T S_w w - 1)$$
$$= w^T S_B w - \lambda \cdot w^T S_w w + \lambda$$

because for a symmetric matrix, $S^T = S$, S_w, S_B are given as scatter matrix.

because for a symmetric matrix, it holds: $w^T S w = \sum_{ij} w_i S_{ij} w_j$

therefore $\nabla_w (w^T S w) = 2 \cdot S w$

therefore $\nabla_w (w^T S_B w) = 2 S_B w$, $\nabla_w (w^T S_w w) = 2 \cdot S_w w$

therefore $\nabla_w L = 2S_B w - 2\lambda S_w w = 0 \Rightarrow S_B w = \lambda S_w w$

$$\frac{\partial L}{\partial \lambda} = -(w^T S_w w - 1) = 0 \Rightarrow w^T S_w w = 1$$

therefore, the solution of the reformulated problem is also a solution of the generalized eigenvalue problem. $S_B w = \lambda S_w w$

(c) as given in task description $S_B = (m_2 - m_1)(m_2 - m_1)^T$

as proved in task 1(b), $S_B w = \lambda S_w w$

$$(m_2 - m_1) \cdot (m_2 - m_1)^T \cdot w = \lambda S_w w$$

because S_w is positive definite, therefore S_w is invertable.

because $(m_2 - m_1)^T \cdot w$ is a number

rewrite: $\alpha = (m_2 - m_1)^T \cdot w$

$$(m_2 - m_1) \cdot \alpha = \lambda \cdot S_w w$$

$$\frac{\alpha}{\lambda} \cdot S_w^{-1} \cdot (m_2 - m_1) = w$$

because as mentioned in task description, we can extend the objective with a constant, and $\frac{\alpha}{\lambda}$ is just a number

~~there for~~ therefore: $w^* = S_w^{-1}(m_2 - m_1)$

□

Task 2 Bounding the error

given: $P(x \mid w_1) = N(\mu, \Sigma)$ $\quad P(error) = \int_x P(error \mid x)\, p(x)\, dx$

$P(x \mid w_2) = N(-\mu, \Sigma)$

$x \in \mathbb{R}^d$

(a) to prove: $P(error \mid x) \leq \sqrt{P(w_1 \mid x)\, P(w_2 \mid x)}$

proof: $P(error \mid x) = \min\{P(w_1 \mid x), P(w_2 \mid x)\}$

note $a := P(w_1 \mid x)$, $b := P(w_2 \mid x)$, $a, b \in [0,1]$

to prove: $\min\{a,b\} \leq \sqrt{ab}$

case 1: $a \leq b$

$a \times a \leq b \times a$

$\Rightarrow \sqrt{a \times a} = a \leq \sqrt{ab}$

case 2: $a > b$

$a \times b \geq b \times b$

$\Rightarrow \sqrt{ab} > \sqrt{b^2} = b$

therefore, it holds: $P(error \mid x) \leq \sqrt{P(w_1 \mid x) \cdot P(w_2 \mid x)}$

(b) to prove: $P(error) \leq \sqrt{P(w_1) \cdot P(w_2)} \cdot \exp\left(\frac{1}{2} \mu^T \Sigma \mu\right)$

$P(error) = \int P(error \mid x) \cdot p(x)\, dx$

in (a) we get $p(error \mid x) \leq \sqrt{P(w_1 \mid x) \cdot P(w_2 \mid x)}$

in addition $P(w_j \mid x) = \dfrac{P(x \mid w_j) \cdot P(w_j)}{P(x)}$

therefore $P(w_1 \mid x) \cdot P(w_2 \mid x) = \dfrac{P(x \mid w_1) \cdot P(w_1) \cdot P(x \mid w_2) \cdot P(w_2)}{P(x) \cdot P(x)}$

$\sqrt{P(w_1 \mid x) \cdot P(w_2 \mid x)} = \sqrt{P(x \mid w_1) \cdot P(w_1) \cdot P(x \mid w_2) \cdot P(w_2)} \cdot \left(\dfrac{1}{P(x)}\right)$

$= \sqrt{P(x \mid w_1) \cdot P(x \mid w_2)} \cdot \sqrt{P(w_1) \cdot P(w_2)} \cdot \left(\dfrac{1}{P(x)}\right)$

$$\Rightarrow \quad P(error) = \int P(error|x) \cdot p(x)\, dx$$

$$\leq \int \sqrt{P(w_1|x) \cdot P(w_2|x)} \cdot P(x) \cdot dx$$

$$= \int \sqrt{P(x|w_1) \cdot P(x|w_2)} \cdot \sqrt{P(w_1) \cdot P(w_2)} \cdot \frac{1}{P(x)} \cdot P(x) \cdot dx$$

$$= \int \sqrt{P(x|w_1) \cdot P(x|w_2)} \cdot \sqrt{P(w_1) \cdot P(w_2)}\, dx$$

because $P(w_1)$, $P(w_2)$ are prior ~~probability~~ ~~prior~~ probability, they are independent from x

$$\Rightarrow \quad P(error) \leq \sqrt{P(w_1) \cdot P(w_2)} \int \sqrt{P(x|w_1) \cdot P(x|w_2)} \cdot dx$$

reminder: to prove: $P(error) \leq \sqrt{P(w_1) \cdot P(w_2)} \cdot \exp(\frac{1}{2} \mu^T \Sigma \mu)$

the formula for the density of a d-dimentional Gaussian:

$$N(\mu, \Sigma)(x) = \frac{1}{(2\pi)^{d/2} \cdot |\Sigma|^{1/2}} \cdot \exp(-\frac{1}{2}(x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu))$$

$\Rightarrow P(x|w_1) \sim N(\mu, \Sigma)$, $P(x|w_2) \sim N(-\mu, \Sigma)$

$$\cancel{P(x|w)} \quad P(x|w_1) = \frac{1}{(2\pi)^{d/2} \cdot |\Sigma|^{1/2}} \cdot \exp(-\frac{1}{2}(x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu))$$

$$P(x|w_2) = \frac{1}{(2\pi)^{d/2} \cdot |\Sigma|^{1/2}} \cdot \exp(-\frac{1}{2}(x+\mu)^T \cdot \Sigma^{-1} \cdot (x+\mu))$$

$$\sqrt{P(x|w_1) \cdot P(x|w_2)} = \frac{1}{(2\pi)^{d/2} \cdot |\Sigma|^{1/2}} \cdot \exp\left(\left[-\frac{1}{2}(x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu)\right] + \left[-\frac{1}{2}(x+\mu)^T \cdot \Sigma^{-1} \cdot (x+\mu)\right]\right)^{\frac{1}{2}}$$

$$= \frac{1}{(2\pi)^{d/2} \cdot |\Sigma|^{1/2}} \cdot \exp(-\frac{1}{4}(x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu) - \frac{1}{4}(x+\mu)^T \cdot \Sigma^{-1} \cdot (x+\mu))$$

because for $a, b \in R^{m \times m}$, $m \in N$, it holds:

$$(a \pm b)^T = a^T \pm b^T$$

therefore

$$(X-u)^T \cdot \Sigma^{-1} \cdot (X-u)$$

$$= (X^T - u^T) \cdot \Sigma^{-1} \cdot (X-u)$$

$$= X^T \Sigma^{-1} X - X^T \Sigma^{-1} u - u^T \Sigma^{-1} X + u^T \Sigma^{-1} u$$

$$(X+u)^T \cdot \Sigma^{-1} \cdot (X+u)$$

$$= X^T \cdot \Sigma^{-1} X + X^T \Sigma^{-1} u + u^T \Sigma^{-1} X + u^T \Sigma^{-1} u$$

$$\Rightarrow \exp\left(-\tfrac{1}{4}(X-u)^T \cdot \Sigma^{-1} \cdot (X-u) - \tfrac{1}{4}(X+u)^T \cdot \Sigma^{-1}(X+u)\right)$$

$$= \exp\left(-\tfrac{1}{2} X^T \Sigma^{-1} \cdot X\right) \cdot \exp\left(-\tfrac{1}{2} u^T \Sigma^{-1} u\right)$$

because $\dfrac{1}{(2\pi)^{d/2} \cdot |\Sigma|^{1/2}} \cdot \exp\left(-\tfrac{1}{2} X^T \cdot \Sigma^{-1} \cdot X\right) = N(0, \Sigma)(X)$

$$\Rightarrow \sqrt{p(X|w_1) \cdot p(X|w_2)} = \exp\left(\tfrac{1}{2} u^T \Sigma^{-1} u\right) \cdot N(0, \Sigma)(X)$$

$$\int \sqrt{p(X|w_1) \cdot p(X|w_2)}\, dx = \exp\left(\tfrac{1}{2} u^T \Sigma^{-1} u\right) \int N(0, \Sigma)(X)$$

$$= \exp\left(\tfrac{1}{2} u^T \Sigma^{-1} u\right) \cdot 1$$

$$= \exp\left(\tfrac{1}{2} u^T \Sigma^{-1} u\right)$$

$$\Rightarrow p(\text{error}) \leq \sqrt{p(w_1) \cdot p(w_2)} \cdot \exp\left(\tfrac{1}{2} u^T \Sigma^{-1} u\right)$$

$\square$

Task 3

given $p(x|w_1) = N\left(\begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}2&0\\0&1\end{bmatrix}\right)$

$P(x|w_2) = N\left(\begin{bmatrix}+1\\+1\end{bmatrix}, \begin{bmatrix}2&0\\0&1\end{bmatrix}\right)$

(a) $\mu_1 = \begin{bmatrix}-1\\-1\end{bmatrix}$ $\mu_2 = \begin{bmatrix}1\\1\end{bmatrix}$ $\Sigma_1 = \Sigma_2 = \Sigma = \begin{bmatrix}2&0\\0&1\end{bmatrix}$

because Fischer's direction is defined up to scaling. it free to take

$S_w = \Sigma = \begin{bmatrix}2&0\\0&1\end{bmatrix}$

$S_w = \Sigma^{-1} = \begin{bmatrix}\frac{1}{2}&0\\0&1\end{bmatrix}$

$\mu_2 - \mu_1 = \begin{bmatrix}2\\2\end{bmatrix}$

$w^* \propto S_w^{-1}(m_2 - m_1) = \begin{bmatrix}\frac{1}{2}&0\\0&1\end{bmatrix} \cdot \begin{bmatrix}2\\2\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix}$

(b)

$J(w) = \dfrac{w^T S_B w}{w^T S_w w}$, $m_1 = \begin{bmatrix}-1\\-1\end{bmatrix}$, $m_2 = \begin{bmatrix}1\\1\end{bmatrix}$, $\Sigma_1 = \Sigma_2 = \Sigma = \begin{bmatrix}2&0\\0&1\end{bmatrix}$

$S_B = (m_2 - m_1)(m_2 - m_1)^T = \left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}-1\\-1\end{bmatrix}\right) \cdot \left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}-1\\-1\end{bmatrix}\right)^T = \begin{bmatrix}2\\2\end{bmatrix} \cdot \begin{bmatrix}2&2\end{bmatrix} = \begin{bmatrix}4&4\\4&4\end{bmatrix}$

$S_w = \Sigma = \begin{bmatrix}2&0\\0&1\end{bmatrix}$

$w = \begin{bmatrix}w_1\\w_2\end{bmatrix} \neq 0$, $(m_2 - m_1) = v = \begin{bmatrix}2\\2\end{bmatrix}$

$w^T \cdot S_B \cdot w = 4(w_1 + w_2)^2$

$w^T \cdot S_w \cdot w = 2w_1^2 + w_2^2$ $\Rightarrow J(w) = \dfrac{4(w_1 + w_2)^2}{2w_1^2 + w_2^2}$, $w \neq 0$

to find $\arg\min(4(w_1 + w_2)^2)$ $\Rightarrow w_1 + w_2 = 0$ $\Rightarrow w_1 = -w_2$

Since $w_1 + w_2 = 0$, then w is orthogonal to $v = \begin{bmatrix}2\\2\end{bmatrix}$

$\Rightarrow w = \alpha\begin{bmatrix}1\\-1\end{bmatrix}$, $\alpha \neq 0$ □