UE 3

Sonntag, 20. Oktober 2024 15:07

a) $\max_{w} \frac{w^T S_B w}{w^T S_W w}$ s.t. $w^T S_W w = 1$

with $S_B := (m_2 - m_1)(m_2 - m_1)^T$

$m_k := \frac{1}{N_k} \sum_{n \in C_k} x_n$

$S_W := S_1 + S_2$

$S_k := \sum_{n \in C_k} (x_n - \mu_k)(x_n - \mu_k)^T$

$\equiv \max_{w} w^T S_B w$ s.t. $w^T S_W w = 1$

Because $w^T S_B w$ is quadratic and convex, maximizing only $w^T S_B w$ without any constraints would yield the trivial solution $\|w\| \to \infty$.
One possibility would be to limit the length of w, e.g., $\|w\| = \frac{1}{\alpha} \Leftrightarrow \|\alpha w\| = 1$
This does not affect the optimization objective because it is invariant to rescaling

$$\frac{w^T S_B w}{w^T S_W w} = \frac{(\alpha w)^T S_B (\alpha w)}{(\alpha w)^T S_W (\alpha w)}$$

However, it is more convenient to constrain $w^T S_W w$, e.g., $w^T S_W w = \frac{1}{\alpha^2} \Leftrightarrow (\alpha w)^T S_W (\alpha w) = 1$ because then the objective simplifies $\frac{w^T S_B w}{w^T S_W w} = w^T S_B w$
Still, enforcing $w^T S_W w = 1$ is equivalent to limiting the length of w.

b) Lagrangian Multipliers:

$$\mathcal{L}(w, \lambda) = w^T S_B w - \lambda (w^T S_W w - 1)$$

$$\frac{\partial \mathcal{L}}{\partial w} = (S_B + S_B^T) w - \lambda (S_W + S_W^T) w$$

$$S_B^T = \left((m_2 - m_1)(m_2 - m_1)^T\right)^T = (m_2 - m_1)^{T^T} (m_2 - m_1)^T = (m_2 - m_1)(m_2 - m_1)^T = S_B \Rightarrow S_B \text{ sym.}$$

$$S_W^T = (S_1 + S_2)^T = S_1^T + S_2^T$$

$$S_k^T = \left(\sum_{n \in C_k} (x_n - \mu_k)(x_n - \mu_k)^T\right)^T = \sum_{n \in C_k} (x_n - \mu_k)^{T^T} (x_n - \mu_k)^T = \sum_n (x_n - \mu_k)(x_n - \mu_k)^T = S_k \Rightarrow S_k \text{ sym.}$$

$$= S_1 + S_2 = S_W \rightarrow S_W \text{ sym.}$$

Used properties:

$\frac{\partial x^T B x}{\partial x} = (B + B^T) x$

$(AB)^T = B^T A^T$

$A^{T^T} = A$

$(A + B)^T = A^T + B^T$

$$= 2 S_B w - 2 \lambda S_W w \overset{!}{=} 0 \quad | : 2$$

$$\Leftrightarrow S_B w = \lambda S_W w$$

$\Rightarrow$ generalized eigval. problem

c) $S_B w = \lambda S_W w \quad | \; S_W^{-1}$

$\Leftrightarrow S_W^{-1} S_B w = \lambda w$

$\Leftrightarrow S_W^{-1} (m_2 - m_1)(m_2 - m_1)^T w = \lambda w \quad | : (m_2 - m_1)^T w \quad (\in \mathbb{R})$

$\Leftrightarrow S_W^{-1} (m_2 - m_1) = \frac{\lambda}{(m_2 - m_1)^T w} w$ rewrite $\lambda' = \frac{\lambda}{(m_2 - m_1)^T w}$

$\Leftrightarrow S_W^{-1} (m_2 - m_1) = \lambda' w$

$\Leftrightarrow w \propto S_W^{-1} (m_2 - m_1)$

$\boxed{2}$ a) From 1st lecture: $P(\text{error} | x) = \min\{P(\omega_1 | x), P(\omega_2 | x)\}$

$\Rightarrow \min\{P(\omega_1 | x), P(\omega_2 | x)\} \leq \sqrt{P(\omega_1 | x) \cdot P(\omega_2 | x)}$

Let $\min\{P(\omega_1 | x), P(\omega_2 | x)\} = p$ (so its either $p = P(\omega_1 | x) < P(\omega_2 | x) = q$ or $p = P(\omega_2 | x) < P(\omega_1 | x) = q$)

and $\max\{P(\omega_1 | x), P(\omega_2 | x)\} = q$

Let $\min\{P(\omega_1|x), P(\omega_2|x)\} = p$ (so its either $p = P(\omega_1|x) < P(\omega_2|x) = q$ or $p = P(\omega_2|x) < P(\omega_1|x) = q$)
and $\max\{P(\omega_1|x), P(\omega_2|x)\} = q$

$$\Leftrightarrow \quad p \leq \sqrt{pq} \quad |^2$$

$$\Leftrightarrow \quad p \cdot p \leq pq \quad | : p \neq 0 \quad \text{otherwise ineq. holds anyway}$$

$$\Leftrightarrow \quad p \leq q$$

$$\Leftrightarrow \quad \min\{P(\omega_1|x), P(\omega_2|x)\} \leq \max\{P(\omega_1|x), P(\omega_2|x)\} \qquad \square$$

(b)

$$P(\text{error}) \overset{\text{marginalization}}{=} \int_x P(\text{error}|x)\, p(x)\, dx$$

$$\overset{(2a)}{\leq} \int_x \sqrt{P(\omega_1|x)\, P(\omega_2|x)}\; p(x)\, dx$$

$$= \int_x \sqrt{P(\omega_1|x)p(x)}\sqrt{P(\omega_2|x)p(x)}\, dx$$

$$\overset{*}{=} \int_x \sqrt{p(x|\omega_1)p(\omega_1)}\sqrt{p(x|\omega_2)p(\omega_2)}\, dx$$

$$* \; \overset{\text{Bayes' theorem}}{p(\omega|x)p(x)} = p(x|\omega)p(\omega)$$

$$= \int_x \sqrt{p(\omega_1)p(\omega_2)}\sqrt{p(x|\omega_1)\, p(x|\omega_2)}\, dx$$

$$= \sqrt{p(\omega_1)p(\omega_2)} \int_x \sqrt{p(x|\omega_1)\, p(x|\omega_2)}\, dx$$

$$p(x|\omega_1)\cdot p(x|\omega_2) = \mathcal{N}(\mu, \Sigma)\cdot\mathcal{N}(-\mu, \Sigma)$$

$$= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}^2}\left(\exp\left\{-\tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}\cdot\exp\left\{-\tfrac{1}{2}(x+\mu)^T\Sigma^{-1}(x+\mu)\right\}\right)$$

$$\exp(a)\cdot\exp(b) = \exp(a+b) \qquad = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}^2}\left(\exp\left\{-\tfrac{1}{2}\left(\left(x^T\Sigma^{-1}x - \cancel{2x^T\Sigma^{-1}\mu} + \mu^T\Sigma^{-1}\mu\right) + \left(x^T\Sigma^{-1}x + \cancel{2x^T\Sigma^{-1}\mu} + \mu^T\Sigma^{-1}\mu\right)\right)\right\}\right)$$

$$= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}^2}\left(\exp\left\{-x^T\Sigma^{-1}x - \mu^T\Sigma^{-1}\mu\right\}\right)$$

$$= \exp\left\{-\mu^T\Sigma^{-1}\mu\right\}\cdot\frac{\exp\{-x^T\Sigma^{-1}x\}}{\sqrt{(2\pi)^d|\Sigma|}^2}$$

$$= \sqrt{p(\omega_1)p(\omega_2)} \int_x \sqrt{\exp\{-\mu^T\Sigma^{-1}\mu\}\,\frac{\exp\{-x^T\Sigma^{-1}x\}}{(2\pi)^d|\Sigma|}}\, dx$$

$$\sqrt{\exp(a)} = \exp\left(\tfrac{a}{2}\right) \qquad = \sqrt{p(\omega_1)p(\omega_2)}\,\exp\left\{-\tfrac{1}{2}\mu^T\Sigma^{-1}\mu\right\}\int_x \frac{\exp\{-\frac{1}{2}x^T\Sigma^{-1}x\}}{\sqrt{(2\pi)^d|\Sigma|}}\, dx$$

$$\int_x p(x)dx = 1 \qquad = \sqrt{p(\omega_1)p(\omega_2)}\,\exp\left\{-\tfrac{1}{2}\mu^T\Sigma^{-1}\mu\right\}\underbrace{\int_x \mathcal{N}(x|0,\Sigma)\, dx}_{=1}$$

$$= \sqrt{p(\omega_1)p(\omega_2)}\,\exp\left\{-\tfrac{1}{2}\mu^T\Sigma^{-1}\mu\right\}$$

$$= \sqrt{p(\omega_1)p(\omega_2)}\, \exp\left\{-\frac{1}{2}\mu^T \Sigma^{-1} \mu\right\} \qquad \overset{\vee}{=} 1$$

$\square$

$\boxed{3}$

(a) from ex. (1c)

$$S_W = (S_1 + S_2) = 2\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

$$S_B = (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T = \left(\begin{pmatrix} 2 \\ 2 \end{pmatrix} (2 \;\; 2) \right) = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}$$

$$w^* = S_W^{-1} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$$

(b) from (1b)

$$S_B w = \lambda S_W w \quad | \; w^T.$$

$$\Leftrightarrow \; w^T S_B w = \lambda \underbrace{w^T S_W w}_{=1}$$

$$\Leftrightarrow \; w^T S_B w = \lambda$$

$\Longrightarrow$ optimization criterion is equal to eigenvalue

$\Longrightarrow$ eigenvector with largest eigenvalue = direction with max. objective

$\Longrightarrow$ eigenvector with smallest eigenvalue = direction with min. objective

from (1b)

$$S_B w = \lambda S_W w$$

$$\Leftrightarrow \; S_W^{-1} S_B w = \lambda w$$

$$\Leftrightarrow \; \begin{pmatrix} 1/4 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix} w = \lambda w$$

$$\Leftrightarrow \; \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} w = \lambda w$$

solve characteristic equation for eigenvals.

$$\det\left(\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = 0 \qquad \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$\Leftrightarrow \; \det\begin{pmatrix} (1-\lambda) & 1 \\ 2 & (2-\lambda) \end{pmatrix} = 0$$

$$\Leftrightarrow (1-\lambda)(2-\lambda) - 2 = 0 \qquad \Longrightarrow \text{ eigval } \lambda_2 = 0$$

$$\Leftrightarrow \cancel{2} - \lambda - 2\lambda + \lambda^2 - \cancel{2} = 0$$

$$\Leftrightarrow \lambda^2 - 3\lambda = 0 \qquad \Longrightarrow \text{ eigval } \lambda_1 = 3$$

eigval $\lambda_1 = 3$ should yield w from (3a)

eigval $\lambda_2 = 0$ should yield w for minimum

$$\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

I $\quad w_1 + w_2 = 0 \quad \Leftrightarrow \quad w_1 = -w_2$

II $\quad 2w_1 + 2w_2 = 0 \quad \Leftrightarrow \quad w_1 = -w_2$

$\rightarrow$ e.g. $\quad w = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is the direction in which the objective is minimized.

For completness

I $\quad w_1 + w_2 = 3w_1 \quad \Rightarrow \quad w_2 = 2w_1$

II $\quad 2w_1 + 2w_2 = 3w_2$

$\rightarrow$ e.g. $\quad w = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$ ✓ is the Fisher direction.

Alternative solution:

$$w^T S_B w = w^T (m_2 - m_1)(m_2 - m_1)^T w$$

Alternative solution:

$$w^T S_B w = w^T (m_2 - m_1)(m_2 - m_1)^T w$$

$$= \left(w^T (m_2 - m_1)\right)^2 \geq 0 \qquad (\text{non-negative})$$

$\Rightarrow$ S_B is positive semi-definite

$\Rightarrow$ find direction v with smallest possible $v^T S_B v$, i.e., $v^T S_B v = 0$

$$v^T S_B v = \left(v^T (m_2 - m_1)\right)^2 \overset{!}{=} 0$$

$$\Leftrightarrow \quad v^T (m_2 - m_1) = 0 \quad \Rightarrow \quad v \text{ is orthogonal to } (m_2 - m_1)$$

$$(m_2 - m_1) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \text{ so } v \propto \begin{pmatrix} 2 \\ -2 \end{pmatrix} \quad \checkmark$$