Ex.1. (a) $\arg\max_{w} \; w^T S_B w \quad \text{s.t. } w^T S_w w = 1$

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

$\Rightarrow \nabla L(w,\lambda) \overset{!}{=} 0 \Leftrightarrow \frac{\partial L}{\partial w} = 2S_B w - 2\lambda S_w w \overset{!}{=} 0 \Leftrightarrow S_B w = \lambda S_w w$

$\frac{\partial L}{\partial \lambda} = 1 - w^T S_w w \overset{!}{=} 0 \qquad w^T S_w w = 1$

Thus the solution to the problem in (a) is also a solution of the problem $S_B w = \lambda S_w w$

(c) Let w^* be a solution of the eigenvalue problem, i.e.:

$S_B w^* = \lambda S_w w^*$

$\Leftrightarrow w^* = S_w^{-1}(m_2 - m_1)\underbrace{\frac{(m_2 - m_1)^T w^*}{\lambda}}_{\text{scalar}}$

(*1 Ex.3. We want to find $\arg\max_{w} \frac{(m_2(w) - m_1(w))^2}{s_2(w) - s_1(w)}$, which by lecture 3 and Ex. 1 is equivalent to the eigenvalue problem

$S_B w = \lambda S_w w$ with the constriction $w^T S_w w = 1$

We have $S_B = (\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T = \left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}-1\\-1\end{bmatrix}\right)\cdot\left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}-1\\-1\end{bmatrix}\right)^T$

$= \begin{bmatrix}2\\2\end{bmatrix}[2 \; 2] = \begin{bmatrix}4 & 4\\4 & 4\end{bmatrix}$

$S_w = S_1 + S_2$ where $S_1 = S_2 = \begin{bmatrix}2 & 0\\0 & 1\end{bmatrix}$ since we know the data generating probabilities $\Rightarrow S_w = \begin{bmatrix}4 & 0\\0 & 2\end{bmatrix} \Rightarrow S_w^{-1} = \begin{bmatrix}\frac{1}{4} & 0\\0 & \frac{1}{2}\end{bmatrix}$

$S_B w = \lambda S_w w \Leftrightarrow S_w^{-1} S_B w = \lambda w \Leftrightarrow \begin{bmatrix}\frac{1}{4} & 0\\0 & \frac{1}{2}\end{bmatrix}\begin{bmatrix}4 & 4\\4 & 4\end{bmatrix} w = \lambda w$

$\begin{bmatrix}1 & 1\\2 & 2\end{bmatrix} w = \lambda w$

$\Rightarrow 0 \overset{!}{=} \det\left(\begin{bmatrix}1-\lambda & 1\\2 & 2-\lambda\end{bmatrix}\right) = (1-\lambda)(2-\lambda) - 2 = 2 - 3\lambda + \lambda^2 - 2 = \lambda(\lambda - 3)$

$\Rightarrow \lambda_1 = 0, \; \lambda_2 = 3$

By $S_B w = \lambda S_w w$ $\lambda = 0$ would result in $S_B w = 0$ thus this solution is not relevant for the Fisher discriminant.

$\Rightarrow \lambda_2 = 3, \quad 0 = \begin{bmatrix}-2 & 1\\2 & -1\end{bmatrix} w \Rightarrow \begin{matrix}-2w_1 + w_2 = 0\\ 2w_1 - w_2 = 0\end{matrix} \Rightarrow w = t\cdot\begin{bmatrix}1\\2\end{bmatrix}$

$1 \overset{!}{=} w^T S_w w = [t \; 2t]\begin{bmatrix}4 & 0\\0 & 2\end{bmatrix}\begin{bmatrix}t\\2t\end{bmatrix} = 4t^2 + 8t^2 = 12t^2$

$\Rightarrow t = \sqrt{\frac{1}{12}} \Rightarrow w = \sqrt{\frac{1}{12}}\begin{bmatrix}1\\2\end{bmatrix}$

*1 For Ex.3 we use the definitions of mu/sk from the lecture.

(b) In order to minimize the ratio $\frac{(m_2(w)-m_1(w))^2}{s_1(w)+s_2(w)}$ we can find a vector s.t. $(m_2(w)-m_1(w))^2=0$, $s_2(w)+s_1(w)=w^T S_W w \neq 0$

We have that $m_2(w)-m_1(w) = w^T(\hat{\mu}_2-\hat{\mu}_1) = w^T \begin{bmatrix}2\\2\end{bmatrix}$

this is zero for $w=\begin{bmatrix}1\\-1\end{bmatrix}$ which also fulfills the constraint

$w^T S_W w = \begin{bmatrix}1 & -1\end{bmatrix}\begin{bmatrix}4 & 0\\0 & 2\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix} = 4+2=6 \neq 0$

Ex. 2. (a) $P(error|x) = \min\{P(\omega_1|x), P(\omega_2|x)\}$

$$= \sqrt{(\min\{P(\omega_1|x), P(\omega_2|x)\})^2}$$

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

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

$$= \sqrt{P(\omega_1)P(\omega_2)} \int_x \sqrt{p(x|\omega_1)\, p(x|\omega_2)}\, \frac{p(x)}{\sqrt{p(x)^2}}\, dx$$

$$= \sqrt{P(\omega_1)P(\omega_2)} \int_x \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\tfrac{1}{4}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) \cdot \exp\left(-\tfrac{1}{4}(x+\mu)^T \Sigma^{-1}(x+\mu)\right) dx$$

$$= \sqrt{P(\omega_1)P(\omega_2)} \int_x \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\tfrac{1}{4}(x-\mu)^T \Sigma^{-1}(x-\mu) - \tfrac{1}{4}(x+\mu)^T \Sigma^{-1}(x+\mu)\right) dx$$

$$= \sqrt{P(\omega_1)P(\omega_2)} \int_x \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\tfrac{1}{4}\left(x^T\Sigma^{-1}x - \mu^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu + x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu\right)\right) dx$$

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

$$= \sqrt{P(\omega_1)P(\omega_2)}\, \exp\left(-\tfrac{1}{2}\mu^T\Sigma^{-1}\mu\right) \int_x \mathcal{N}(0,\Sigma)\, dx = \sqrt{P(\omega_1)P(\omega_2)}\, \exp\left(-\tfrac{1}{2}\mu^T\Sigma^{-1}\mu\right)$$