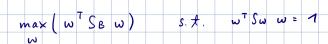
Exercise 1: Fisher Discriminant (10+10+10 P)

The objective function to find the Fisher Discriminant has the form

$$\max_{oldsymbol{w}} rac{oldsymbol{w}^{ op} oldsymbol{S}_B oldsymbol{w}}{oldsymbol{w}^{ op} oldsymbol{S}_W oldsymbol{w}}$$

where $S_B = (m_2 - m_1)(m_2 - m_1)^{\top}$ is the between-class scatter matrix and S_W is within-class scatter matrix, assumed to be positive definite. Because there are infinitely many solutions (multiplying \boldsymbol{w} by a scalar doesn't change the objective), we can extend the objective with a constraint, e.g. that enforces $\boldsymbol{w}^{\top} \boldsymbol{S}_{W} \boldsymbol{w} = 1.$

(a) Reformulate the problem above as an optimization problem with a quadratic objective and a quadratic



(b) Show using the method of Lagrange multipliers that the solution of the reformulated problem is also a solution of the generalized eigenvalue problem:

$$S_B w = \lambda S_W w$$

$$f(\omega) = \omega^{T} S_{B} \omega$$
 $g(\omega) = \omega^{T} S_{W} \omega - 1$

$$\mathcal{L}(\omega, \alpha) = \omega^{T} S_{B} \omega + \alpha (\omega^{T} S_{W} \omega - 1), \text{ where } \alpha \text{ is a constant}$$

$$\nabla \omega \mathcal{L}(\omega, \alpha) = 2 S_{B} \omega + 2 \alpha S_{W} \omega = 0$$

$$\Rightarrow S_{8}\omega = - \propto S_{\omega}\omega$$

$$\Rightarrow S_{8}\omega = \lambda S_{\omega}\omega$$

(c) Show that the solution of this optimization problem is equivalent (up to a scaling factor) to

$$m{w}^{\star} = m{S}_W^{-1}(m{m}_1 - m{m}_2)$$

$$(\Rightarrow) \omega = \frac{1}{\lambda} \cdot S\omega^{-1} \cdot SB \cdot \omega$$

$$(\Rightarrow) \omega = \frac{1}{\lambda} \cdot S\omega^{-1} (m_{\Sigma} - m_{1}) (m_{\Sigma} - m_{1})^{T} \omega$$

= w = Sw (m2-m1) Scalar

Exercise 2: Bounding the Error (10+10 P)

The direction learned by the Fisher discriminant is equivalent to that of an optimal classifier when the class-conditioned data densities are Gaussian with same covariance. In this particular setting, we can derive a bound on the classification error which gives us insight into the effect of the mean and covariance parameters on the error.

Consider two data generating distributions $P(\boldsymbol{x}|\omega_1) = \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and $P(\boldsymbol{x}|\omega_2) = \mathcal{N}(-\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{x} \in \mathbb{R}^d$. Recall that the Bayes error rate is given by:

$$P(\text{error}) = \int_{\boldsymbol{x}} P(\text{error}|\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}$$

(a) Show that the conditional error can be upper-bounded as:

$$P(\text{error}|\boldsymbol{x}) \leq \sqrt{P(\omega_1|\boldsymbol{x})P(\omega_2|\boldsymbol{x})}$$

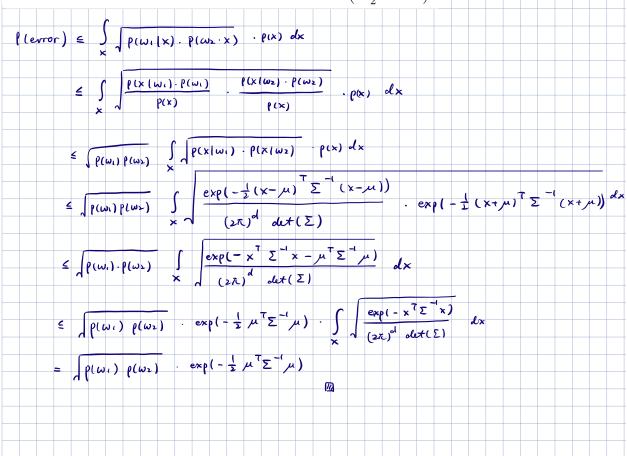
$$P(\omega_{1}(x)) = \min \left\{ P(\omega_{1}(x)), P(\omega_{2}(x)) \right\} = M_{-\infty} \left(P(\omega_{1}(x)), P(\omega_{2}(x)) \right)$$

$$M_{0} \left(P(\omega_{1}(x)), P(\omega_{2}(x)) \right) = \prod_{i=1}^{2} P(\omega_{i}(x)) = \prod_{i=1}^{2} P(\omega_{i}(x)), P(\omega_{2}(x))$$

$$\Rightarrow \min \left\{ P(\omega_{1}(x)), P(\omega_{2}(x)) \right\} \leq \prod_{i=1}^{2} P(\omega_{1}(x)), P(\omega_{2}(x))$$

(b) Show that the Bayes error rate can then be upper-bounded by:

$$P(\text{error}) \leq \sqrt{P(\omega_1)P(\omega_2)} \cdot \exp\left(-\frac{1}{2}\boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)$$

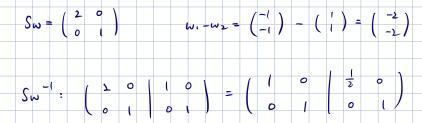


Exercise 3: Fisher Discriminant (10 + 10 P)

Consider the case of two classes ω_1 and ω_2 with associated data generating probabilities

$$p(\boldsymbol{x}|\omega_1) = \mathcal{N}\left(\begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right) \quad \text{and} \quad p(\boldsymbol{x}|\omega_2) = \mathcal{N}\left(\begin{pmatrix} +1 \\ +1 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

(a) Find for this dataset the Fisher discriminant \boldsymbol{w} (i.e. the projection $y = \boldsymbol{w}^{\top} \boldsymbol{x}$ under which the ratio between inter-class and intra-class variability is maximized).



$$\omega = S\omega^{-1}(\omega_1 - \omega_2) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

(b) Find a projection for which the ratio is minimized.

$$\omega = \left(\begin{array}{c} -1 \\ 1 \end{array} \right)$$