Machine Learning I at TU Berlin

Assignment 6 - Group PTHGL

November 26, 2018

Exercise 1

Fisher Discriminant

(a)

$$\max_{\omega} J(\omega)$$

$$\frac{\partial J(\omega)}{\partial \omega} = \frac{(\frac{d}{d\omega}\omega^T S_B \omega)\omega^T S_w \omega - (\frac{d}{d\omega}\omega^T S_w \omega)\omega^T S_B \omega}{(\omega^T S_w \omega)^2} \qquad \frac{d}{dx}(a^T X a) = (X + X^T)$$

S_w and S_B symmetric. ($X = X^T$)

$$= \frac{(2 S_B \omega)\omega^T S_w \omega - (2 S_w \omega)\omega^T S_B \omega}{(\omega^T S_w \omega)^2} = 0$$

$$= (S_B \omega)\omega^T S_w \omega - (S_w \omega)\omega^T S_B \omega = 0 \qquad | : (\omega^T S_w \omega)$$

$$= S_B \omega - S_w \omega \underbrace{\frac{\omega^T S_B \omega}{\omega^T S_B \omega}}_{\lambda} = 0$$

$$S_B \omega = \lambda S_w \omega$$

(b)

S_w is a symmetric,positive semi-definite and non singular matrix, so S_w is invertible.

$$S_B \omega = \lambda S_w \omega$$

$${S_w}^{-1} S_B \omega = \lambda \omega$$

S_B is always in the direction of $(m_1 - m_2)$:

$$S_B \omega (M_1 - m_2) \underbrace{(m_1 - m_2)^T \omega}_{\alpha} = \alpha (m_1 - m_2)$$

$$\longrightarrow {S_w}^{-1} \alpha (m_1 - m_2) = \lambda \omega$$

We just need the normalized vector of ω. For this reason α and λ should both be 1.

$$\omega = {S_w}^{-1}(m_1 - m_2)$$

Exercise 2

LDA vs. Optimal Classification

(a)

Given is the following joint probability distributions:

$$p(\boldsymbol{x}|\omega_1) = \frac{1}{16} \cdot 1_{0 \leq x_1 \leq 4} \cdot 1_{-1 \leq x_2 \leq 3} \qquad p(\boldsymbol{x}|\omega_2) = \frac{1}{16} \cdot 1_{-4 \leq x_1 \leq 0} \cdot 1_{-3 \leq x_2 \leq 1}$$

To derive the mean and covariances, we need the marginal probability distributions:

$$\begin{aligned} p(x_1|\omega_1) &= \int_{-1}^{3} p(\boldsymbol{x}|\omega_1) dx_2 & p(x_1|\omega_2) &= \int_{0}^{4} p(\boldsymbol{x}|\omega_1) dx_2 \\ &= \frac{1}{4} \cdot 1_{0 \leq x_1 \leq 4} & &= \frac{1}{4} \cdot 1_{-1 \leq x_2 \leq 3} \\ p(x_2|\omega_1) &= \int_{-3}^{1} p(\boldsymbol{x}|\omega_1) dx_1 & p(x_2|\omega_2) &= \int_{-4}^{0} p(\boldsymbol{x}|\omega_1) dx_1 \\ &= \frac{1}{4} \cdot 1_{-4 \leq x_1 \leq 0} & &= \frac{1}{4} \cdot 1_{-3 \leq x_2 \leq 1} \end{aligned}$$

With $\mu = E[\mathbf{x}] = \int \mathbf{x} p(\mathbf{x}) dx$ we get the means:

$$\mu_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

For the covariance matrices we only need to calculate the variance of one marginal distributions, because the uniform distributions $p(x_1|\omega_1), p(x_2|\omega_1) p(x_1|\omega_2)$ and $p(x_2|\omega_2)$ all have the same interval range, and therefore the same variance. Furthermore, since in x_1 and x_2 $p(\boldsymbol{x}|\omega_1)$ and $p(\boldsymbol{x}|\omega_1)$ are orthogonal, the off-diagonal terms have to be equal to 0. With this we get the following covariance matrices:

$$\Sigma_1 = \begin{bmatrix} \frac{4}{3} & 0 \\ 0 & \frac{4}{3} \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} \frac{4}{3} & 0 \\ 0 & \frac{4}{3} \end{bmatrix}$$

As expected, the covariances are equal, which is also a condition for LDA. Therefore we define $\Sigma = \Sigma_1 = \Sigma_2$

(b)

$$\begin{aligned}\mathbf{w} &= \Sigma^{-1}[\mu_{\mathbf{1}} - \mu_{\mathbf{2}}] \\ &= \begin{bmatrix} 3/4 & 0 \\ 0 & 3/4 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ \frac{3}{2} \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\mathbf{b} &= \frac{1}{2}\left[T + \begin{bmatrix} -2 & -1 \end{bmatrix} \begin{bmatrix} 3/4 & 0 \\ 0 & 3/4 \end{bmatrix} \begin{bmatrix} -2 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 3/4 & 0 \\ 0 & 3/4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right] \\ &= \frac{1}{2}T\end{aligned}$$

(d)

Bayesian error rate $E = 1 - \sum_{i=1}^{L} \int_{W_i} P(\omega_i)p(x|\omega_i)dx$, L number of class
$P(\omega_1) = P(\omega_2) = 0.5$

$$\begin{aligned}E &= 1 - 0.5 * \frac{1}{16}\left(\int_{-1}^{3}\int_{0}^{4} 1dx_1dx_2 + \int_{-3}^{1}\int_{-4}^{0} 1dx_1dx_2\right) \\ &= 1 - \frac{1}{32}\left(\int_{-1}^{3} 4dx_2 + \int_{-3}^{1} 4dx_2\right) \\ &= 0\end{aligned}$$

(f)

An alternative linear discriminant can be:
  if $x_1 > 0$ then $\mathbf{x} \in \omega_1$
  if $x_1 < 0$ then $\mathbf{x} \in \omega_2$