Exercise 1 - Rewriting the Fisher criterion for LDA

Let

$$J(w) = \frac{\langle w, \mu_+ - \mu_- \rangle^2}{\sigma^2_{w,+} + \sigma^2_{w,-}} = \frac{N}{D}$$

First let us evaluate the numerator.

$$\begin{aligned}
N &= \langle w, \mu_+ - \mu_- \rangle^2 \\
&= \langle w, \mu_+ - \mu_- \rangle \langle w, \mu_+ - \mu_- \rangle \\
&= \langle w, \mu_+ - \mu_- \rangle \langle \mu_+ - \mu_-, w \rangle \qquad \text{Since } \langle a, b \rangle = \langle b, a \rangle \\
&= w^{,T} (\mu_+ - \mu_-)(\mu_+ - \mu_-)^T w \\
&= w^T C_B w \\
&= \langle w, C_B w \rangle
\end{aligned}$$

Moving on the the denominator.

$$\begin{aligned}
\sigma^2_{w,+} &= \frac{1}{m_+} \sum_{i:Y_i=+1} \left(\langle w, X_i - \mu_+ \rangle^2 \right) \\
&= \frac{1}{m_+} \sum_{i:Y_i=+1} \left(w^T (X_i - \mu_+)(X_i - \mu_+)^T w \right) \qquad \text{See evaluation for } N \\
&= w^T \left(\frac{1}{m_+} \sum_{i:Y_i=+1} (X_i - \mu_+)(X_i - \mu_+)^T \right) w \\
\sim \sigma^2_{w,-} &= w^T \left(\frac{1}{m_-} \sum_{i:Y_i=-1} (X_i - \mu_-)(X_i - \mu_-)^T \right) w \\
\implies D = \sigma^2_{w,+} + \sigma^2_{w,-} &= w^T \left[\left\{ \frac{1}{m_+} \sum_{i:Y_i=+1} (X_i - \mu_+)(X_i - \mu_+)^T \right\} + \left\{ \frac{1}{m_-} \sum_{i:Y_i=-1} (X_i - \mu_-)(X_i - \mu_-)^T \right\} \right] w \\
&= w^T C_W w \\
&= \langle w, C_W w \rangle
\end{aligned}$$

Therefore,

$$J(w) = \frac{\langle w, \mu_+ - \mu_- \rangle^2}{\sigma^2_{w,+} + \sigma^2_{w,-}} = \frac{\langle w, C_B w \rangle}{\langle w, C_W w \rangle}$$

Exercise 4 - Complexity of multiclass classification

Part (a) - Learning complexity $= c\,(m_1^2 + m_2^2)$

- One-vs-All: $\binom{k}{1} \times c \left[\left(\frac{1}{k} n \right)^2 + \left\{ \left(1 - \frac{1}{k} \right) n \right\}^2 \right] = \frac{cn^2}{k} \left[(k-1)^2 + 1 \right]$
- One-vs-One: $\binom{k}{2} \times c \left[\left(\frac{1}{k} n \right)^2 + \left(\frac{1}{k} n \right)^2 \right] = \frac{cn^2}{k} (k-1)$

Part (b) - Learning complexity $= c\,(m_1 + m_2)$

- One-vs-All: $\binom{k}{1} \times c \left[\left(\frac{1}{k} n \right) + \left(1 - \frac{1}{k} \right) n \right] = cnk$
- One-vs-One: $\binom{k}{2} \times c \left[\left(\frac{1}{k} n \right) + \left(\frac{1}{k} n \right) \right] = cn\,(k-1)$

Comments

- One-vs-One is faster in both cases.
- For scaling n, the algorithm for part (b) is preferable.

Exercise 5 - Parameter selection by the training error

Overfitting