Exercise 1 - Rewriting the Fisher criterion for LDA

Let

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

First let us evaluate the numerator.

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

Moving on the the denominator.

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

Therefore,

$$J\left(w\right)=\frac{\left\langle w,\mu_{+}-\mu_{-}\right\rangle^{2}}{\sigma_{w,+}^{2}+\sigma_{w,-}^{2}}=\frac{\left\langle w,C_{B}w\right\rangle}{\left\langle w,C_{W}w\right\rangle}$$

Exercise 4 - Complexity of multiclass classification

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

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

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

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

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