Tutorial 5

Machine Learning and Big Data for Economics and Finance

List of activities

I. Complete **Section 4.6 Lab: Logistic Regression, LDA, QDA, and KNN**, subsections 4.6.3, 4.6.4 and 4.6.6.

II. Complete the list of exercises in this tutorial.

Exercise 1. Linear discriminant analysis

Derive formula 4.24 in the textbook.

Solution.

$$
\begin{aligned}
\frac{p_1(x)}{p_2(x)} &= \frac{\dfrac{\pi_1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_1)^2}}{\sum_{k'} \pi_{k'} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_{k'})^2}}}{\dfrac{\pi_2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_2)^2}}{\sum_{k'} \pi_{k'} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_{k'})^2}}} \\
&= \frac{\pi_1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_1)^2}}{\pi_2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_2)^2}} \\
&= \frac{\pi_1 e^{-\frac{1}{2\sigma^2}(x-\mu_1)^2}}{\pi_2 e^{-\frac{1}{2\sigma^2}(x-\mu_2)^2}}
\end{aligned}
$$

This implies that

$$
\begin{aligned}
\log\left(\frac{p_1(x)}{p_2(x)}\right) &= \log\left(\frac{\pi_1}{\pi_2}\right) - \frac{1}{2\sigma^2}(x-\mu_1)^2 + \frac{1}{2\sigma^2}(x-\mu_2)^2 \\
&= \log\left(\frac{\pi_1}{\pi_2}\right) - \frac{1}{2\sigma^2}x^2 + x\frac{\mu_1}{\sigma^2} - \frac{\mu_1^2}{2\sigma^2} + \frac{1}{2\sigma^2}x^2 - x\frac{\mu_2}{\sigma^2} + \frac{\mu_2^2}{2\sigma^2} \\
&= \log\left(\frac{\pi_1}{\pi_2}\right) + \frac{\mu_2^2 - \mu_1^2}{2\sigma^2} + x\left(\frac{\mu_1 - \mu_2}{\sigma^2}\right)
\end{aligned}
$$

Thus the log odds is linear with $c_0 = \log\left(\frac{\pi_1}{\pi_2}\right) + \frac{\mu_2^2 - \mu_1^2}{2\sigma^2}$ and $c_1 = \frac{\mu_1 - \mu_2}{\sigma^2}$.