Business Analytics
Exercise Sheet 7

Martin Wistuba (wistuba@ismll.de)
Information Systems and Machine Learning Lab (ISMLL)
Universität Hildesheim

18 June 2014
Submission until 24 June 2014 23:59

Exercise 19: Discriminant Analysis (5 Points)

Scientists compared the earth of Iowa which contains a specific bacterium (class 1) with other earth that does not contain it (class 2). They observed the variables x_1 (pH value) and x_2 (nitrogen content). The number of instances pro class, the mean of the vectors and the covariance matrix for both kind of earths is given as follows:

$$n_1 = 13, \qquad n_2 = 10$$
$$\mu_1 = \begin{pmatrix} 7.8 \\ 45 \end{pmatrix}, \qquad \mu_2 = \begin{pmatrix} 5.9 \\ 20.8 \end{pmatrix}$$
$$\mathbf{\Sigma}_{W1} = \begin{pmatrix} 0.5 & 4.5 \\ 4.5 & 147.2 \end{pmatrix}, \quad \mathbf{\Sigma}_{W2} = \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 24.2 \end{pmatrix}$$

a) Estimate the discriminant functions for both classes.

b) Assign the observation $x = \begin{pmatrix} 6 & 52.5 \end{pmatrix}^T$ to one of the both classes.

c) Is this a linear or a quadratic discriminant analysis? Mention differences between LDA and QDA.

x	y	x	y
-4	4	3	3
-4	3	4	3
-3	3	4	4
-4	-3	3	-3
-4	-4	4	-4
-3	-3	4	-3
3	3	4	-4

Tabelle 1: Data to be clustered.

Exercise 20: K-Means Clustering (5 points)

(a) Explain the optimization function of the K-Means clustering in your own words.

(b) What is the primary difference between K-Means and K-Medoids?

(c) Apply the K-Means Clustering for two iterations on the data in Table 1 for $k = 4$. The first cluster center was randomly chosen to be $(3, 3)$.

Submission

- Electronically to wistuba@ismll.de. Text submitted as pdf, code submitted as source files. No archives.