MS1b: SDM - Problem Sheet 4

1. (LDA) Let W be the $p \times p$-dimensional within-class covariance matrix estimated from 2-class data (X_i, Y_i), with $X_i = (X_i^{(1)}, \dots, X_i^{(p)})$ a p-dimensional predictor variable and $Y_i \in \{1, 2\}$, given by

$$W = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_{Y_i})^T(X_i - \hat{\mu}_{Y_i}),$$

where $\hat{\mu}_k$ is the mean across all observations with class k.

   (a) Explain how to transform the data so that the within-class covariance is equal the identity.

   (b) Denote by B the estimated between class covariance matrix of the transformed data and by $\hat{\mu}_1$ and $\hat{\mu}_2$ the estimated class means of the transformed data. For K classes

$$B = \frac{1}{K}\sum_{i=1}^{K}(\hat{\mu}_i - \hat{\mu})^T(\hat{\mu}_i - \hat{\mu})$$

   with $\hat{\mu}$ the mean vector of all observations $X_i, i = 1, \dots, n$ and $n\hat{\mu} = n_1\hat{\mu}_1 + n_2\hat{\mu}_2$, where n_k are the number of observations in class k. Observe that B can be written as an outer product of $(\hat{\mu}_1 - \hat{\mu}_2)$ with itself and hence show that $(\hat{\mu}_1 - \hat{\mu}_2)$ is a eigenvector of B with eigenvalue λ_1 say.

   (c) Calculate the p eigenvalues and verify that λ_1 is the largest. Interpret this.

2. (Decision Theory, Naïve Bayes) Suppose there are two classes $Y \in \{1, 2\}$ and that given the class, $Y = k$, data vector X has two independent Bernoulli components $X = (X^{(1)}, X^{(2)})$, so that for $Y_i = k$, for both $j = 1, 2$,

$$X^{(j)} \sim \text{Bernoulli}(p_{kj}).$$

Suppose the four probability parameters $p_{ki},\ i = 1, 2,\ k = 1, 2$ are known. Show that the Bayes classifier $\hat{Y}(x)$ minimizing the total Risk for the $0 - 1$ loss function has a linear discriminant function of the form

$$\hat{Y}(x) = \arg\min_{k=1,2}\quad \alpha_k + \beta_k^T x.$$

and find the values of α_k, β_k.

3. (LDA, QDA) Suppose we have a two-class setup with classes -1 and 1, that is $\mathcal{Y} = \{-1, 1\}$ and a 2-dimensional predictor variable $X = (X^{(1)}, X^{(2)})$. We find that the means of the two groups are at $\hat{\mu}_{-1} = (-1, -1)$ and $\hat{\mu}_1 = (1, 1)$ respectively. The a priori probabilities are equal.

(a) Applying LDA, the covariance matrix is estimated to be, for some value of $0 \leq \rho \leq 1$,

$$\hat{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Find the decision boundary as a function of ρ.

(b) Now applying QDA, we estimate the covariance matrices for group -1 as

$$\hat{\Sigma}_{-1} = \begin{pmatrix} 5 & 0 \\ 0 & 1/5 \end{pmatrix},$$

and for group 1 as

$$\hat{\Sigma}_1 = \begin{pmatrix} 1/5 & 0 \\ 0 & 5 \end{pmatrix}.$$

Find now the decision boundary and draw a sketch of it in the two-dimensional plane.