# 9

## MVA

(2012 – 2013 Academic Year: Tutorial Questions)

**Exercises**

9.1 Compute the coefficients of a Fisher discriminant to separate samples of data $A$ and $B$ given the means $\mu$ and covariance matrices $\sigma^2$ below.

$$\mu_A = \begin{pmatrix} 1.0 \\ 2.0 \end{pmatrix}, \quad \mu_B = \begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix}, \quad \sigma^2_A = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}, \quad \sigma^2_B = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

9.2 Compute the coefficients of a Fisher discriminant to separate samples of data $A$ and $B$ given the means $\mu$ and covariance matrices $\sigma^2$ below.

$$\mu_A = \begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix}, \quad \mu_B = \begin{pmatrix} -1.0 \\ 2.0 \end{pmatrix}, \quad \sigma^2_A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \sigma^2_B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

9.3 Compute the coefficients of a Fisher discriminant to separate samples of data $A$ and $B$ given the means $\mu$ and covariance matrices $\sigma^2$ below.

$$\mu_A = \begin{pmatrix} 0.2 \\ 0.7 \\ 0.1 \end{pmatrix}, \qquad \mu_B = \begin{pmatrix} 0.6 \\ 1.5 \\ 0.2 \end{pmatrix},$$

$$\sigma^2_A = \begin{pmatrix} 0.1 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.2 \end{pmatrix}, \quad \sigma^2_B = \begin{pmatrix} 0.15 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix}.$$

9.4 Compute the coefficients of a Fisher discriminant to separate samples of data $A$ and $B$ given the means $\mu$ and covariance matrices $\sigma^2$ below.

$$\mu_A = \begin{pmatrix} 1.0 \\ 0.7 \\ 0.5 \end{pmatrix}, \qquad \mu_B = \begin{pmatrix} 0.6 \\ 1.0 \\ 0.2 \end{pmatrix},$$

$$\sigma^2_A = \begin{pmatrix} 0.2 & 0.1 & 0.0 \\ 0.1 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.2 \end{pmatrix}, \quad \sigma^2_B = \begin{pmatrix} 0.15 & 0.0 & 0.1 \\ 0.0 & 0.3 & 0.0 \\ 0.1 & 0.0 & 0.3 \end{pmatrix}.$$

9.5 The ratio of momentum and energy ($p/E$) for a charged particle can be used to distinguish between different species (for example $e$, $\mu$, and $\pi$) and has a value between zero and one. Define a Bayesian classifier based on Gaussian distribtions with $\mu = 1.0$ for an electron, 0.5

for a $\pi$ meson and 0.0 for a $\mu$, all three PDFs having a width $\sigma = 0.2$, and classify the following events: $p/E = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$.

9.6 A data sample $\Omega$ of events is comprised of signal and background components. The signal is distributed according to a Gaussian PDF with $\mu = 0.0$ and $\sigma = 1.0$, whereas the background is uniform. Use a Bayesian classifier to categorise the events: $\Omega = \{-3, -2, 0, 1, 4, 5\}$.

9.7 Use a Bayesian classifier to categorise events for the following problem. Two signals are distributed in an $(x, y)$ space, with two-dimensional unit Gaussian distributions centred on $(0, 0)$ and $(1, 2)$, for types $A$ and $B$, respectively. Classify the following events according to this model:

| i | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $x_i$ | $-1.0$ | 0.0 | 1.0 | 3.0 | 2.0 | 0.5 |
| $y_i$ | 1.0 | 2.0 | 1.0 | 3.0 | $-1.0$ | 1.0 |

9.8 Use a Bayesian classifier to categorise events from the previous question using the following model: type $A$ are distributed uniformly, and type $B$ are distributed according to a two-dimensional unit Gaussian centered at $(1, 2)$.

9.9 Signal events are distributed according to $P_A = 3(2x - x^2)/4$, and background events are uniform ($P_B = 1/2$) in the interval $x \in [0, 2]$. It is expected that there are 10 signal and 20 background events in the data. Optimise the cut value $X_c$ with regard to the test statistic signal/background in the interval $[0, X_c]$.

9.10 Do the values for the signal and background yields given in the previous question play a role in the optimal cut value obtained?

9.11 Compute the separation between two distributions of events where $\mu_A = 1$, $\sigma_A = 2$ and $\mu_B = 3$, $\sigma_B = 1$.

9.12 Compute the separation between two distributions of events where $\mu_A = 1$, $\sigma_A = 1$ and $\mu_B = 0$, $\sigma_B = 1$.

9.13 Signal events are distributed according to $P_A = 6(x - x^2)$, and background events are uniform ($P_B = 1$) in the interval $x \in [0, 1]$. Optimise the cut value $X_c$ with regard to the test statistic signal/background in the interval $[0, X_c]$.

9.14 Signal events are distributed according to $P_A = 12(x^2 - x^3)$, and background events are uniform ($P_B = 1$) in the interval $x \in [0, 1]$. Optimise the cut value $X_c$ with regard to the test statistic signal/background in the interval $[0, X_c]$.

9.15 How can one improve the signal to background significance found in the previous questions?