Machine Intelligence

TUTORIAL SHEET 4: Solutions

1. Mahalanobis distance is a special kind of metric for measuring the distance of a point $\mathbf{x}$ from the mean of a distribution with covariance matrix Σ. Under the assumption that two class distributions have the same covariance matrix, the Mahalanobis distance between the respective means of these two distributions measures the separability of the two classes. It is defined as

$$(\underline{\mu_1} - \underline{\mu_2})^T \Sigma^{-1} (\underline{\mu_1} - \underline{\mu_2})$$

where Σ is the common covariance matrix. In contrast the Euclidean metric is defined as

$$(\underline{\mu_1} - \underline{\mu_2})^T (\underline{\mu_1} - \underline{\mu_2})$$

2. Sequential forward selection algorithm is an example of suboptimal feature selection algorithms which can be used when optimal search is computationally prohibitive. It is a bottom up algorithm which starts from the individually best feature or best feature pair. After initialisation, at each stage of the algorithm one selects as the next feature the one that in combination with those already selected gives the best performance. In other words, if n features are already selected, then choose as the $(n+1)st$ the feature which gives the maximum value of the criterion function when combined with the first n features.

   (a) Starting from the best pair $\{x_1, x_4\}$ we have three possible sets to examine: $\{x_1, x_2, x_4\}$, $\{x_1, x_3, x_4\}$ and $\{x_1, x_4, x_5\}$ Using the matrix inversions supplied with the question we can readily compute the Mahalanobis distance in all three cases. For instance for the set $\{x_1, x_2, x_4\}$ we get

$$\Delta\underline{\mu} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1.5 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0.5 \\ 1 \end{bmatrix}$$

   and

$$J(x_1, x_2, x_4) = \Delta\underline{\mu}^T \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix} \Delta\underline{\mu} = 7$$

$$J(x_1, x_3, x_4) = \frac{83}{13}$$

   and

$$J(x_1, x_4, x_5) = 4.68$$

   Thus the best triplet, in the sense of the Mahalanobis distance is $\{x_1, x_2, x_4\}$.

   (b) Note that the Euclidean distance will pick the triplet $\{x_1, x_4, x_5\}$ for which the (squared) Euclidean distance of 6 is maximum. The Euclidean distance ignores the second order statistics of the distributions when selecting the best set of features.

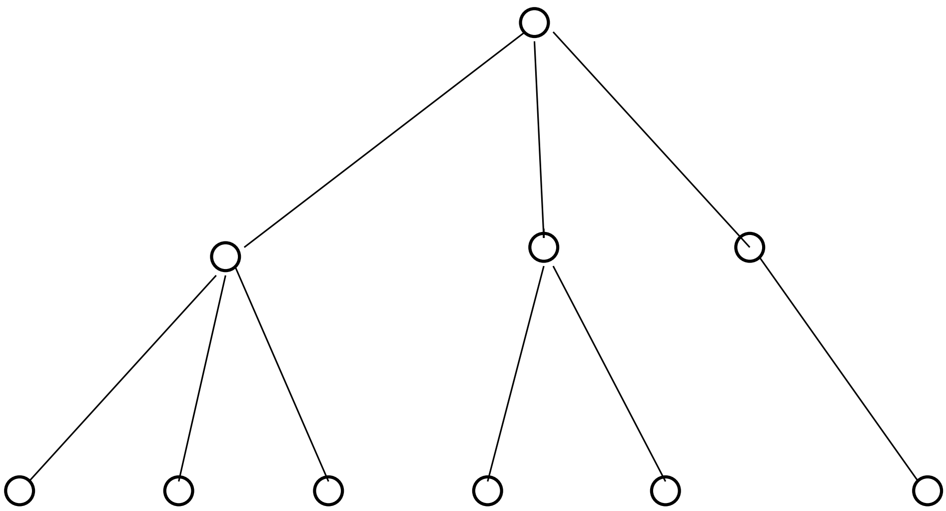

Figure 1: Search tree structure

3. (a) Selection of nodes for the first level of the tree:

Drop x_1

$$\mu = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$|\Sigma| = 15 \qquad \Sigma^{-1} = \begin{bmatrix} \frac{3}{5} & \frac{1}{5} & 0 \\ \frac{1}{5} & \frac{2}{5} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$J = \begin{bmatrix} 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{1}{5} & 0 \\ \frac{1}{5} & \frac{2}{5} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \frac{25}{3}$$

Drop x_2

$$\mu = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 4 & -3 & 0 \\ -3 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$|\Sigma| = 9 \qquad \Sigma^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{4}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$J = \frac{95}{3}$$

Drop x_3

$$\mu = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$|\Sigma| = 21 \qquad \Sigma^{-1} = \begin{bmatrix} \frac{2}{7} & \frac{-1}{7} & 0 \\ \frac{-1}{7} & \frac{4}{7} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$J = \frac{160}{21}$$

Drop x_4

$$\mu = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 4 & 1 & -3 \\ 1 & 2 & -1 \\ -3 & -1 & 3 \end{bmatrix}$$

$$|\Sigma| = 5 \qquad \Sigma^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & \frac{3}{5} & \frac{1}{5} \\ 1 & \frac{1}{5} & \frac{7}{5} \end{bmatrix}$$

$$J = 28$$

The measurements should be allocated to nodes at level 1 in the ascending order of the criterion magnitude so as to maximise the chances of node rejection.

Exploring the right most branch of the tree by droping x_4 and x_2 we have

$$\mu = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 4 & -3 \\ -3 & 3 \end{bmatrix} \quad |\Sigma| = 3 \quad \Sigma^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & \frac{4}{3} \end{bmatrix}$$

$$J = \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \frac{79}{3}$$

Since all the other nodes at level 1 have criterion value less than $\frac{79}{3}$, there is no need to explore the tree further. The optimal solution found is set $X = \{x_1, x_3\}$ with $J(X) = \frac{79}{3}$.

(b) If we use the Euclidean metric instead of the Mahalanobis distance then this will be equivalent to assuming that the covariance matrix is diagonal. The optimal set of features can be then selected on the basis of their individual merit. Accordingly, we have

$$J(x_1) = (\mu_{11} - \mu_{21})^2 = 9 \quad J(x_2) = 1 \quad J(x_3) = 4 \quad J(x_4) = 16$$

Thus the optimal set selected using the Euclidean distance will contain features x_1 and x_4. By comparison with the features selected using the Mahalanobis metric it is apparent that the feature pair $X = \{x_1, x_4\}$ will not be the most discriminative set.