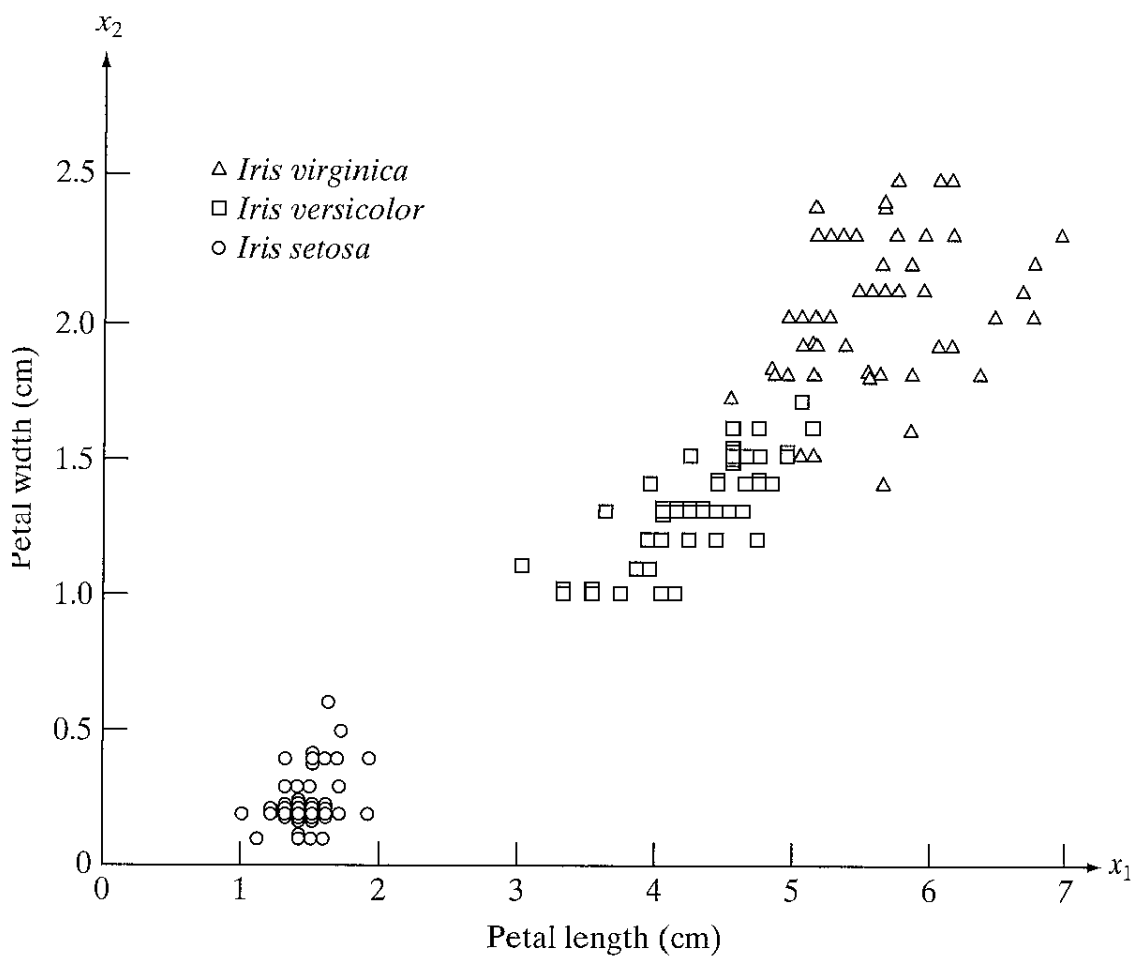


## Solution of selected exercises

### Exercises INF 4300 related to the lecture 30.10.13

#### 2. Finding the decision functions for a minimum distance classifier.

A classifier that uses diagonal covariance matrices is often called a minimum distance classifier, because a pattern is classified to class that is closest when distance is computed using Euclidean distance.



- In the above figure, find the class means just by looking at the plot.
- If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

**Problem 12.1**

(a) By inspection, the mean vectors of the three classes are, approximately,  $\mathbf{m}_1 = (1.5, 0.3)^T$ ,  $\mathbf{m}_2 = (4.3, 1.3)^T$ , and  $\mathbf{m}_3 = (5.5, 2.1)^T$  for the classes Iris setosa, versicolor, and virginica, respectively. The decision functions are of the form given in Eq. (12.2-5). Substituting the above values of mean vectors gives:

$$d_1(x) = x^T \mathbf{m}_1 - \frac{1}{2} \mathbf{m}_1^T \mathbf{m}_1 = 1.5x_1 + 0.3x_2 - 1.2$$

$$d_2(x) = x^T \mathbf{m}_2 - \frac{1}{2} \mathbf{m}_2^T \mathbf{m}_2 = 4.3x_1 + 1.3x_2 - 10.1$$

$$d_3(x) = x^T \mathbf{m}_3 - \frac{1}{2} \mathbf{m}_3^T \mathbf{m}_3 = 5.5x_1 + 2.1x_2 - 17.3$$

(b) The decision boundaries are given by the equations

$$d_{12}(x) = d_1(x) - d_2(x) = -2.8x_1 - 1.0x_2 + 8.9 = 0$$

$$d_{13}(x) = d_1(x) - d_3(x) = -4.0x_1 - 1.8x_2 + 16.1 = 0$$

$$d_{23}(x) = d_2(x) - d_3(x) = -1.2x_1 - 0.8x_2 + 7.2 = 0$$

A plot of these boundaries is shown in Fig. P12.1.

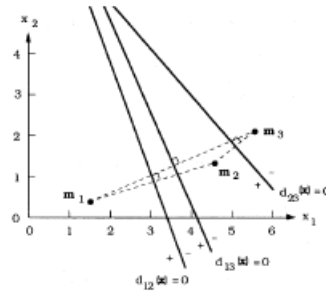


Figure P12.1

### 3. Discriminant functions

A classifier that uses Euclidean distance computes distance from pattern  $\mathbf{x}$  to class  $j$  as:

$$D_j(x) = \|\mathbf{x} - \mu_j\|$$

Show that classification with this rule is equivalent to using the discriminant function

$$d_j(x) = x^T \mu_j - \frac{1}{2} \mu_j^T \mu_j$$

Solution:

**Problem 12.2**

From the definition of the Euclidean distance,

$$D_j(x) = \|\mathbf{x} - \mathbf{m}_j\| = [(\mathbf{x} - \mathbf{m}_j)^T (\mathbf{x} - \mathbf{m}_j)]^{1/2}$$

Since  $D_j(x)$  is non-negative, choosing the smallest  $D_j(x)$  is the same as choosing the smallest  $D_j^2(x)$ , where

$$\begin{aligned} D_j^2(x) &= \|\mathbf{x} - \mathbf{m}_j\|^2 = (\mathbf{x} - \mathbf{m}_j)^T (\mathbf{x} - \mathbf{m}_j) \\ &= \mathbf{x}^T \mathbf{x} - 2\mathbf{x}^T \mathbf{m}_j + \mathbf{m}_j^T \mathbf{m}_j \\ &= \mathbf{x}^T \mathbf{x} - 2 \left( \mathbf{x}^T \mathbf{m}_j - \frac{1}{2} \mathbf{m}_j^T \mathbf{m}_j \right) \end{aligned}$$

We note that the term  $\mathbf{x}^T \mathbf{x}$  is independent of  $j$  (that is, it is a constant with respect to  $j$  in  $D_j^2(x)$ ,  $j = 1, 2, \dots$ ). Thus, choosing the minimum of  $D_j^2(x)$  is equivalent to choosing the maximum of  $(\mathbf{x}^T \mathbf{m}_j - \frac{1}{2} \mathbf{m}_j^T \mathbf{m}_j)$ .

❖ Example:

Given  $\omega_1, \omega_2 : P(\omega_1) = P(\omega_2)$  and  $p(\underline{x}|\omega_1) = N(\underline{\mu}_1, \Sigma)$ ,

$$p(\underline{x}|\omega_2) = N(\underline{\mu}_2, \Sigma), \underline{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \underline{\mu}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$$

classify the vector  $\underline{x} = \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix}$  using Bayesian classification :

- $\Sigma^{-1} = \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix}$

- Compute Mahalanobis  $d_m$  from  $\mu_1, \mu_2 : d^2_{m,1} = [1.0, 2.2]$

$$\Sigma^{-1} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952, d^2_{m,2} = [-2.0, -0.8] \Sigma^{-1} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$$

- Classify  $\underline{x} \rightarrow \omega_1$ . Observe that  $d_{E,2} < d_{E,1}$