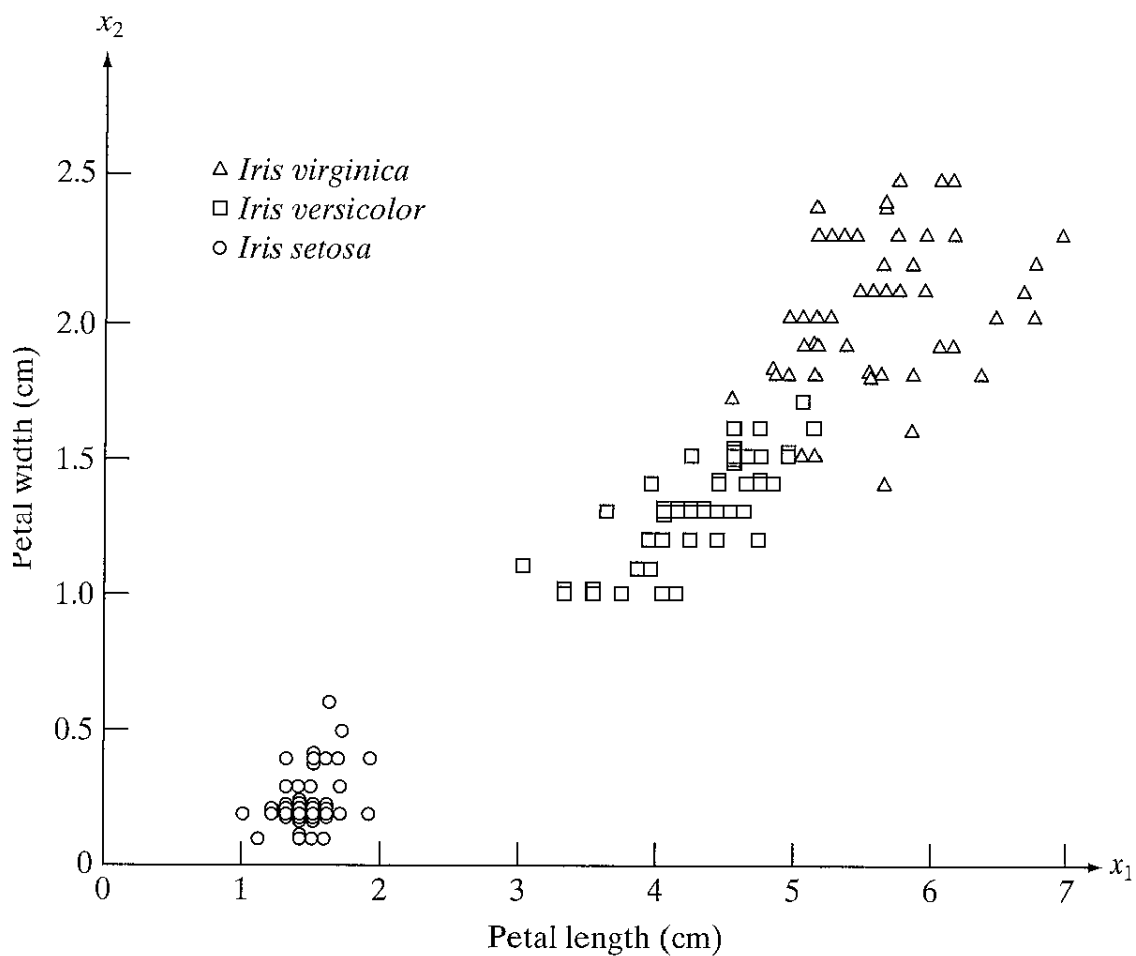


Solution of selected exercises

Exercises IN 5520 related to the lecture 12.10.21

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:

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

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

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

(b) The decision boundaries are given by the equations

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

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

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

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

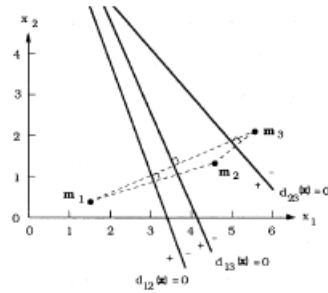


Figure P12.1

3. Discriminant functions

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

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

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

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

Solution:

Problem 12.2

From the definition of the Euclidean distance,

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

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

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

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 $D_j^2(\mathbf{x})$, $j = 1, 2, \dots$). Thus, choosing the minimum of $D_j^2(\mathbf{x})$ 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 $\mu_1, \mu_2 : d_{m,1}^2 = [1.0, 2.2]$
 $\Sigma^{-1} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952, d_{m,2}^2 = [-2.0, -0.8] \Sigma^{-1} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$
- Classify $\underline{x} \rightarrow \omega_1$. Observe that $d_{E,2} < d_{E,1}$

Exercise 3: Classification (From Exam 2015)

- a) Consider a two-dimensional feature vector and a set of points in 2D feature space: (-3,6) (-2,4) (-1, 2) (0,0) (1, -2) (2, -4) (3, -6)

Show that the covariance matrix between the two features is :

$$\Sigma = \begin{bmatrix} 4 & -8 \\ -8 & 16 \end{bmatrix}$$

Show all your calculations.

Answer:

$$\mu_1 = (-3 - 2 - 1 + 0 + 1 + 2 + 3) / 7 = 0$$

$$\mu_2 = (6 + 4 + 2 + 0 + 2 + 4 + 7) / 7 = 0$$

$$\sigma_{12} = \frac{1}{N} \sum_i (x_1 - \mu_1)(x_2 - \mu_2) =$$

$$\frac{1}{7} [(-3-0)(6-0) + (-2-0)(4-0) + (-1-0)(2-0) + (0-0)(0-0) + (-1)(2) + (-2)(4) + (-3)(6)] = -56/7 = -8$$

$$\sigma_1^2 = \frac{1}{N} \sum_i (x_1 - \mu_1)^2 = \frac{1}{7} [3^2 + 2^2 + 1^2 + 0^2 + 1^2 + 2^2 + 3^2] = 28/7 = 4$$

$$\sigma_2^2 = \frac{1}{N} \sum_i (x_2 - \mu_2)^2 = \frac{1}{7} [6^2 + 4^2 + 2^2 + 0^2 + 2^2 + 4^2 + 7^2] = 16$$

- b) Given the two features defined above, would you base your classification on 1 or 2 features? Justify your answer.

Answer: We note that the points line on a straight line, thus the two features are linear dependent, and there is no need to use more than one of them (and the 2D covariance matrix is singular).

- c) The discriminant functions for a multivariate Gaussian classifier are given as:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i)$$

Consider two classes with equal prior probability and

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Sigma_1^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \quad \Sigma_2^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \quad |\Sigma_1| = 1 \quad |\Sigma_2| = 1$$

Can the discriminant function be simplified in this case?

Answer: Somewhat,, we can avoid the $\ln 2\pi$ and $\ln P(\omega_i)$ and also the determinant of the covariance matrices, which is equal.

- d) Classify the point $\mathbf{x} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ by computing the value of the discriminant functions and

assign it to the class corresponding to the highest probability.

Answer:

$$g_1(\mathbf{x}) = -\frac{1}{2} [3-1, 0-0]^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3-1 \\ 0 \end{bmatrix} - 0.5 \ln 1 = -2$$

...

$$g_2(\mathbf{x}) = -\frac{1}{2} [3-0, 0-2]^T \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} - 0.5 \ln 1 = -5/2$$

Here we have ignore the two class-independent terms. We should assign it to the class with maximum value, so the pattern will be assigned to class 1.

- e) Explain how classifier sensitivity and specificity are computed, and discuss their importance on a medical classification problem.

Answer: Very short: Sensitivity: $TP/(TP+FN)$ The probability that the test is positive given that the patient is sick. Specificity: $TN/(TN+FP)$. The probability that the test is negative given that the patient is not sick. For a sick/healthy scenario we want all sick to be diagnosed as sick, so sensitivity is important and we can allow a lower specificity.

- f) Let us assume that we have a 2-class classification problem with a 1-dimensional feature vector $f(x)$ which is exponentially distributed given the class-conditional parameter λ_i :

$$f(x) = \lambda_i \exp^{-\lambda_i x}$$

Find an expression for the decision boundary for this classification problem.

Answer:

The decision boundary is the point T where

$$\lambda_1 e^{-\lambda_1 T} = \lambda_2 e^{-\lambda_2 T} \Leftrightarrow$$

$$\frac{\lambda_1}{\lambda_2} = e^{(\lambda_1 - \lambda_2)T} \Leftrightarrow$$

$$T = \frac{\ln\left(\frac{\lambda_1}{\lambda_2}\right)}{\lambda_1 - \lambda_2}$$

Exercise 4: Classification (From Exam 2016)

Consider a two-dimensional feature vector and a set of points from 2 classes in 2D feature space:

Class 1 has points: (3,0), (5,0), (7,0), (5,2), (5, -2)

Class 2 has points: (0,5), (0,3), (0,7), (-2, 5), (2,5)

The discriminant function for Gaussian classifier is in the general form:

$$g(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i)$$

- a) Compute the mean for each class

Solution:

$$\mu_1 = \begin{bmatrix} 5 \\ 0 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

- b) Show that the covariance between feature 1 and 2 is 0 for both classes.

For class 1: $\sigma_{1,12} = 1/5[(3-5)(0-0) + (5-5)(0-0) + (7-5)(0-0) + (5-5)(2-0) + (5-5)(-2-0)] = 0$

For class 2: $\sigma_{2,12} = 1/5[(0-0)(5-5) + 0(0-0)(3-5) + (0-0)(7-5) + (-2-0)(5-5) + (2-0)(5-5)] = 0$

- c) Show how the discriminant function can be simplified in this case with a classifier with equal diagonal covariance matrices.

Can be simplified to

$$g_i(x) = \mu_{iT} x - \frac{1}{2} \mu_{iT} \mu_i + \ln P(\omega_i)$$

- d) Find an expression for the decision boundary using this simplified discriminant function.

Can be simplified to

$$g_i(x) = g_j(x) \Leftrightarrow \mu_{iT} x - \frac{1}{2} \mu_{iT} \mu_i + \ln P(\omega_i) = \mu_{jT} x - \frac{1}{2} \mu_{jT} \mu_j + \ln P(\omega_j)$$

- e) Compute the decision boundary if we assume equal prior probabilities.

The equation is:

$$5x_1 - 25 = 5x_2 - 25 \Leftrightarrow x_2 = x_1$$

- f) Sketch the class means and the decision boundary in a plot if we assume that the two classes have equal prior probability.
- g) If $P(\omega_1) = 0.75$, in which direction will the decision boundary change? Indicate this on the plot.