MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(w_j|x) = \frac{p(x|w_j)\,P(w_j)}{p(x)}} \quad (1)$$

- Any probability can be used to model $p(x|w_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|w_j) = \frac{1}{(2\pi)^{n/2}\,|\Sigma_j|^{\frac{1}{2}}}\, e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(w_i|\bar{x}) = \frac{p(\bar{x}|w_j)\,P(w_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

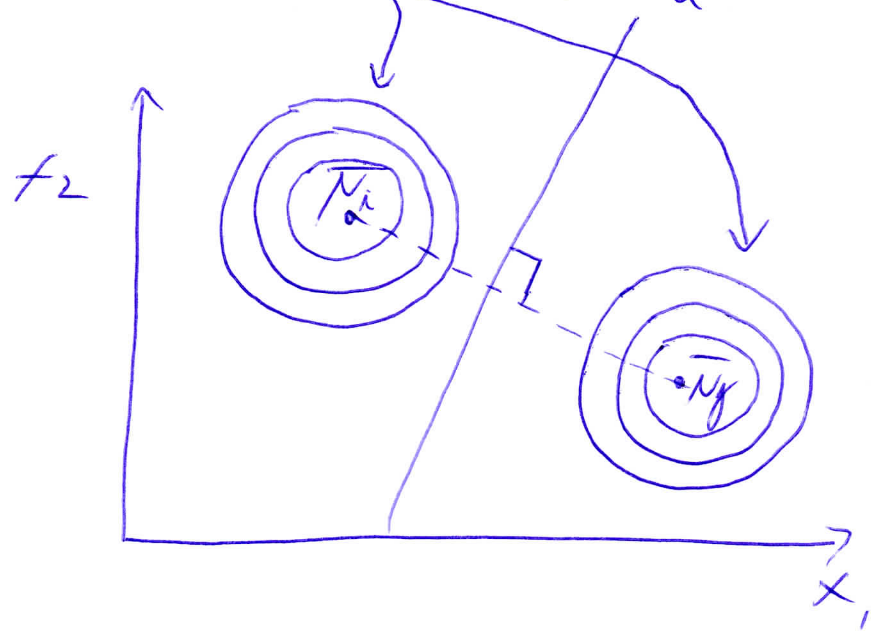
We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \quad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the Euclidean distance.

This is known as the minimum distance classifier.

Euclidean distance



Case 2: Common covariance matrix

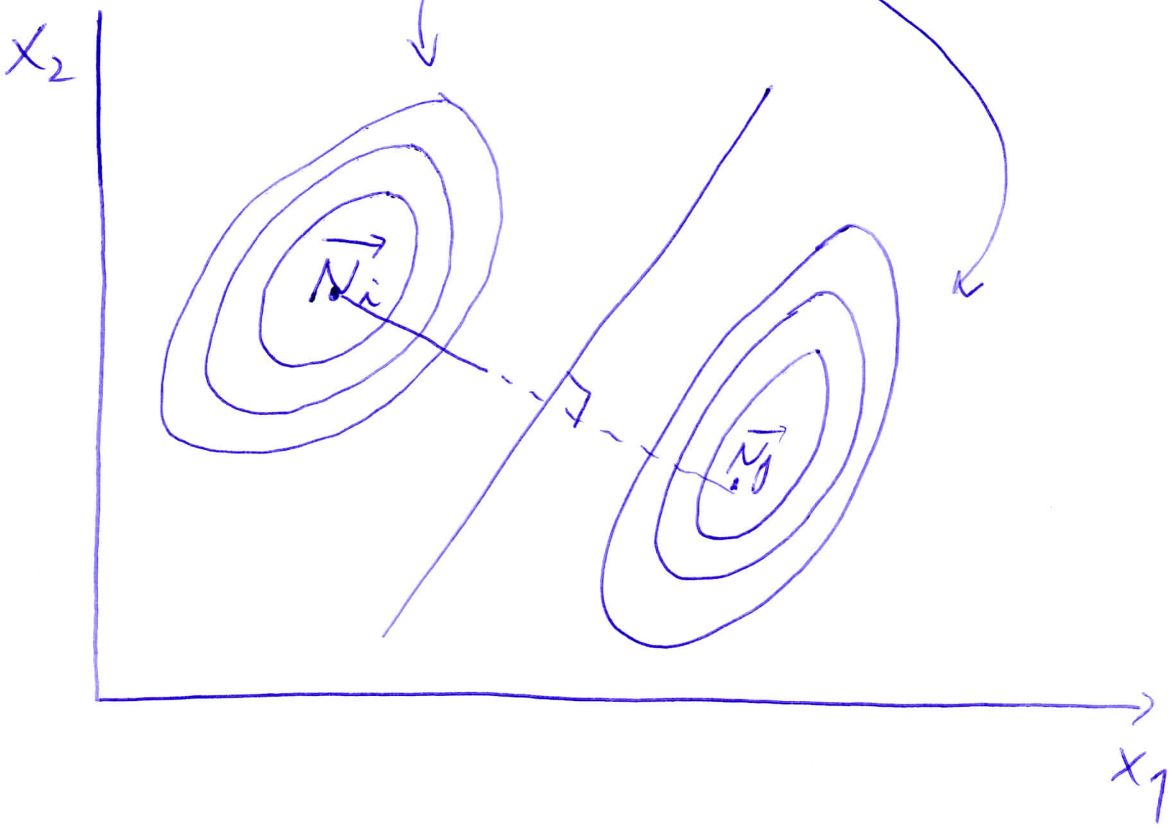
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j)}$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

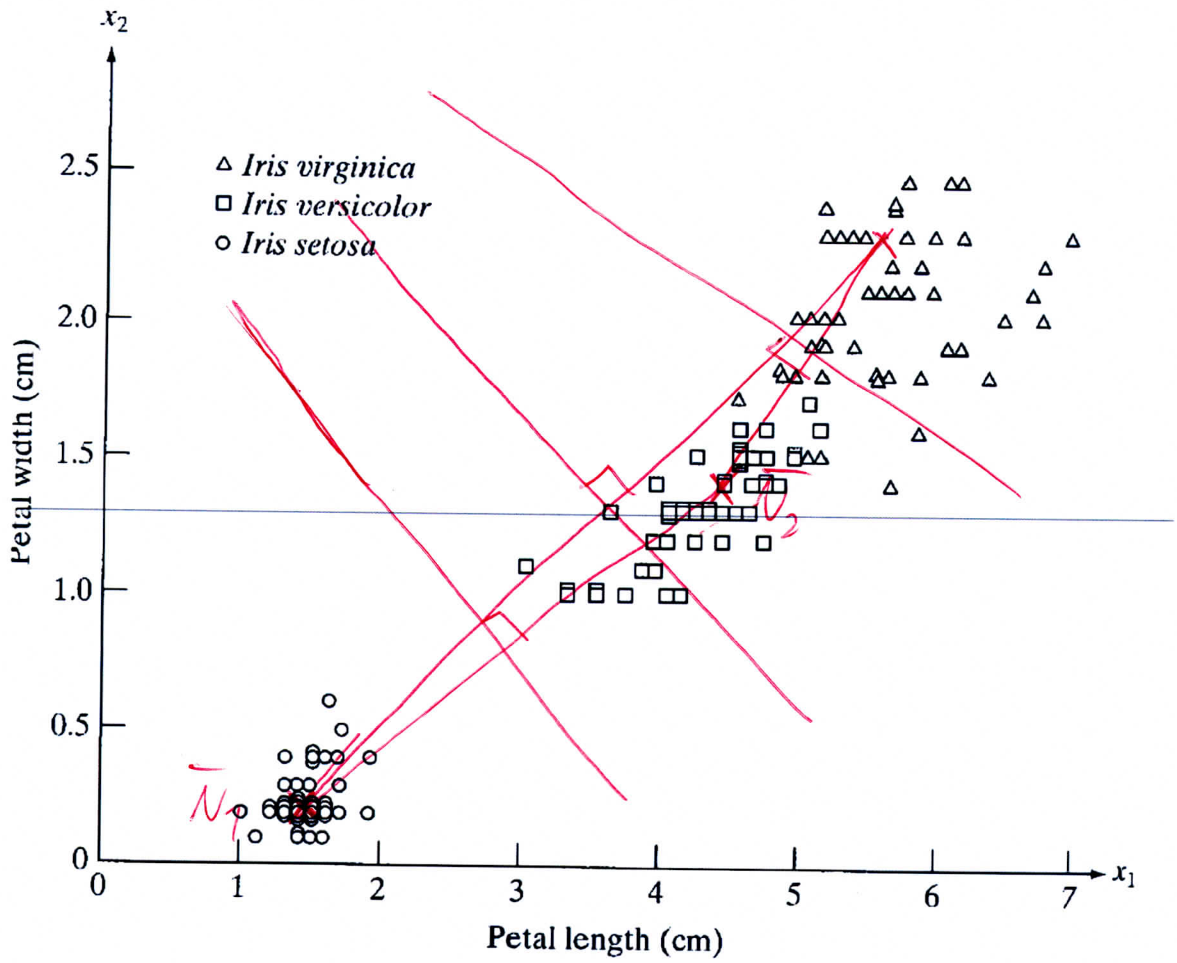
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

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

So we get:

$$G_j^2(x) = \|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2\left(\bar{x}^T\bar{\mu}_j + \tfrac{1}{2}\mu_j^T\mu_j\right)$$

$\bar{x}^T\bar{x}$ is independent of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$, ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} [1,5 \;\; 1,4] \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2} [1,07 \;\; 0,54] \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0,1205$

$g_3(\bar{x}) = -4,7095$

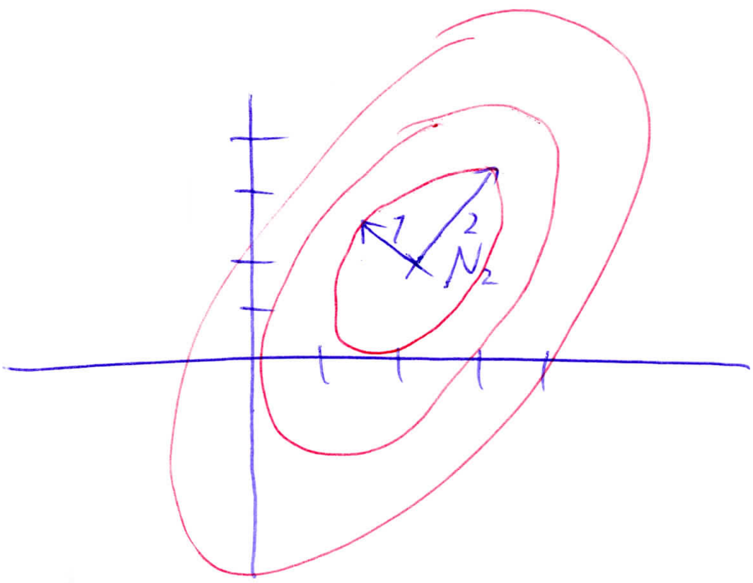
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0,895 \\ 0,447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0,447 \\ 0,895 \end{bmatrix}$

$\lambda_1 = 1 \quad , \quad \lambda_2 = 2$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}$, $\Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2,2 \end{bmatrix}\begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}\begin{bmatrix} 1 \\ 2,2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0,62 & 1,06 \end{bmatrix}\begin{bmatrix} 1 \\ 2,2 \end{bmatrix}$$

$$= -1,4760$$

$$g_2(\bar{x}) = -1,836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)\,P(\omega_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}\,|\Sigma_j|^{\frac{1}{2}}}\, e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)\,P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x} \mid \omega_j) P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln\left(p(\bar{x} \mid \omega_j) P(\omega_j)\right) = \ln\left(p(\bar{x} \mid \omega_j)\right) + \ln\left(P(\omega_j)\right).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x} - \bar{\mu}_j) - \frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

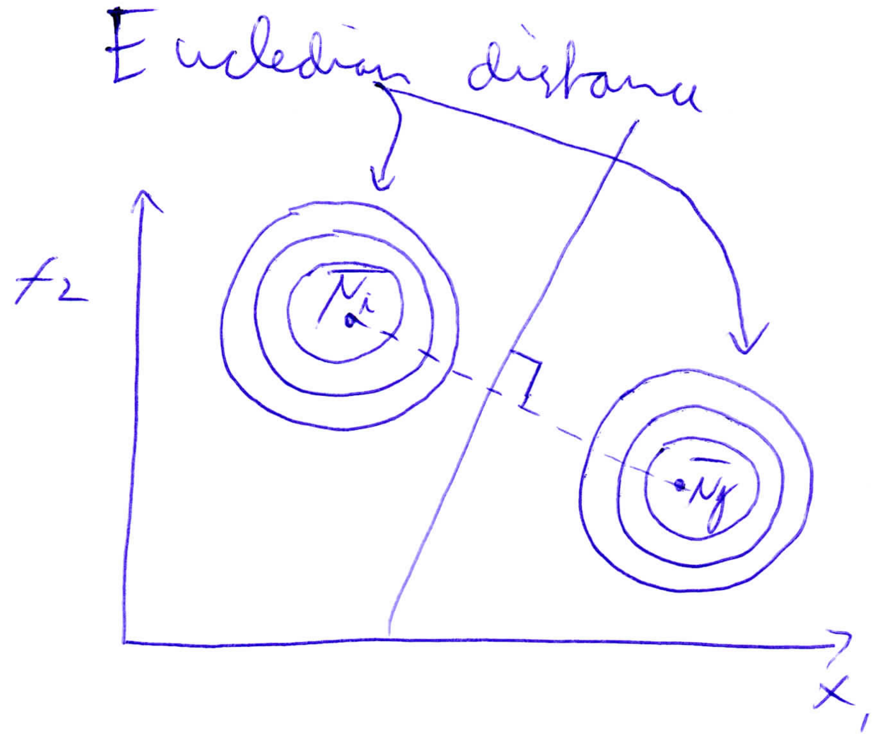
$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \qquad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the Euclidean distance.

This is known as the minimum distance classifier.


Eucledian distance
$\bar{\mu}_i$
$\bar{\mu}_j$
x_2
x_1

Case 2: Common covariance matrix

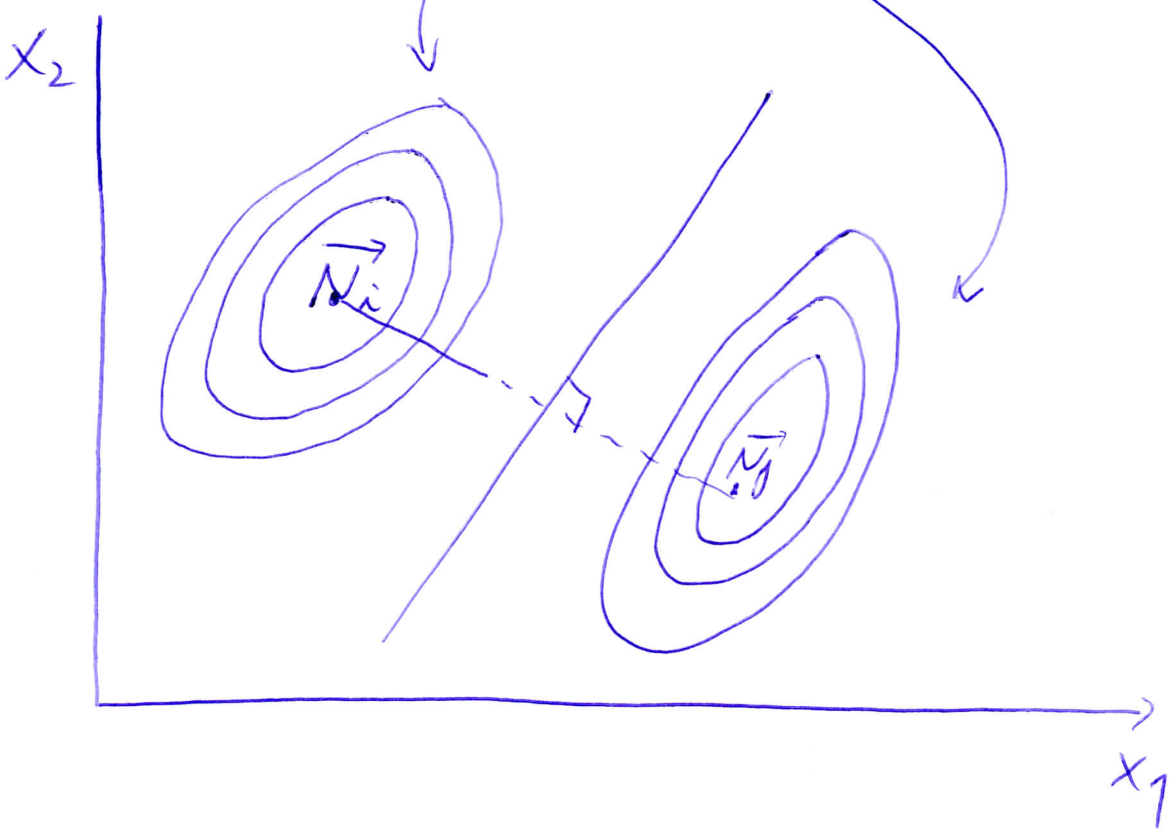
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

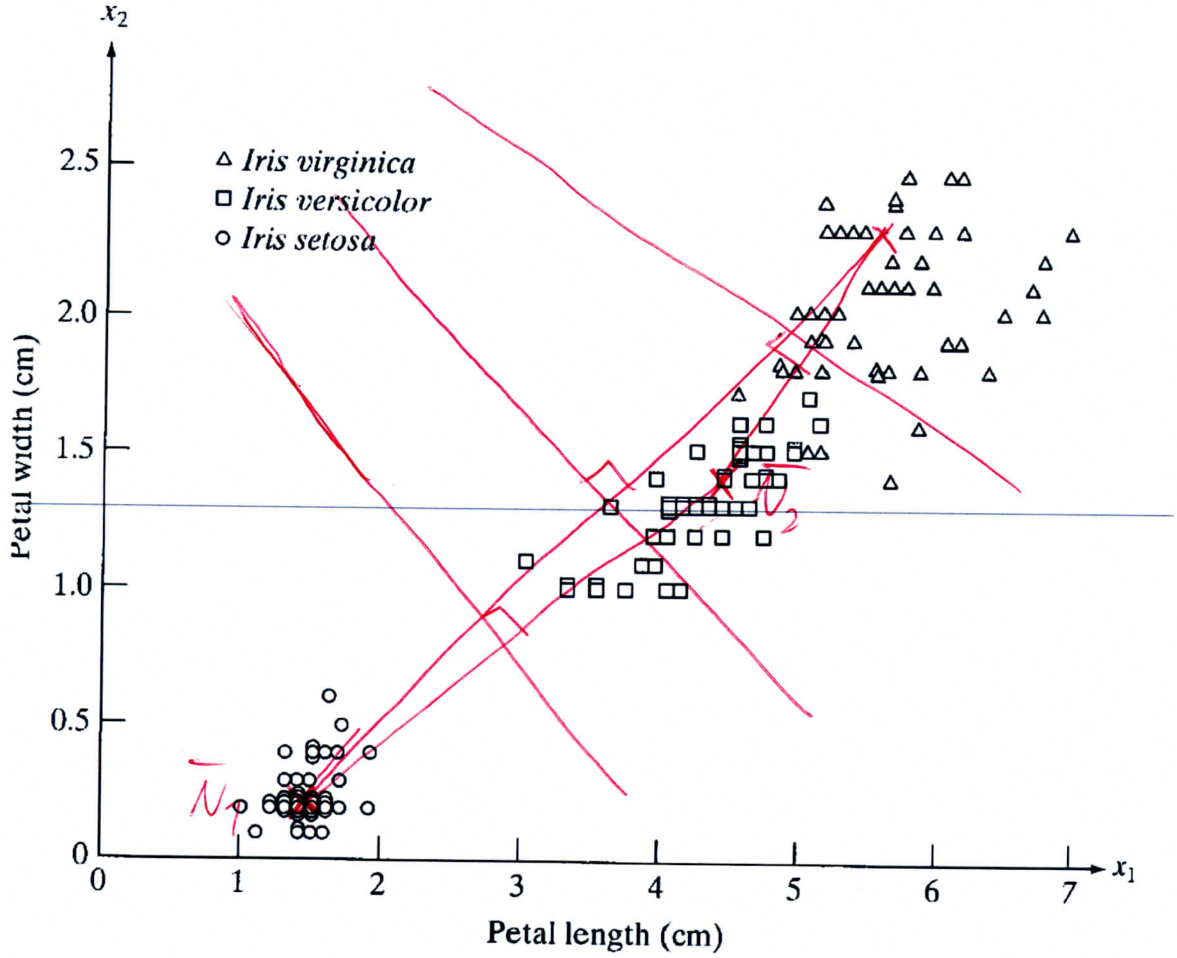
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T(\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2\left(\bar{x}^T\bar{\mu}_j + \tfrac{1}{2}\mu_j^T\mu_j\right)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2}\begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2}[1,5 \;\; 1,4]\begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2}[1,07 \;\; 0,54]\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0,1205$

$g_3(\bar{x}) = -4,7095$

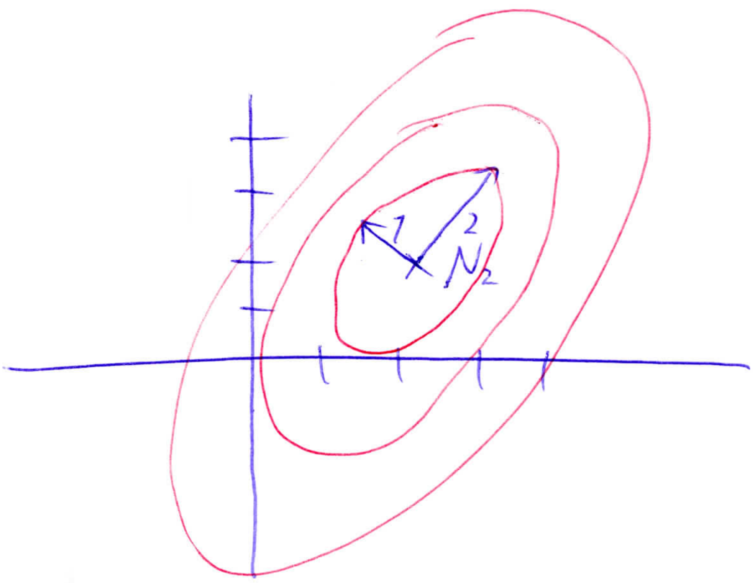
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0,895 \\ 0,447 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0,447 \\ 0,895 \end{bmatrix}$

$\lambda_1 = 1$, $\lambda_2 = 2$



Ex 5) We have

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

$$\overline{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \overline{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix}$, this gives us

$$g_1(\overline{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2} \begin{bmatrix} 1 & 2,2 \end{bmatrix} \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} 0,62 & 1,06 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix}$$

$$= -1,4760$$

$$g_2(\overline{x}) = -1,836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(w_j|x) = \frac{p(x|w_j)P(w_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|w_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|w_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}. \text{ And}$$

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(w_i|\bar{x}) = \frac{p(\bar{x}|w_j)P(w_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

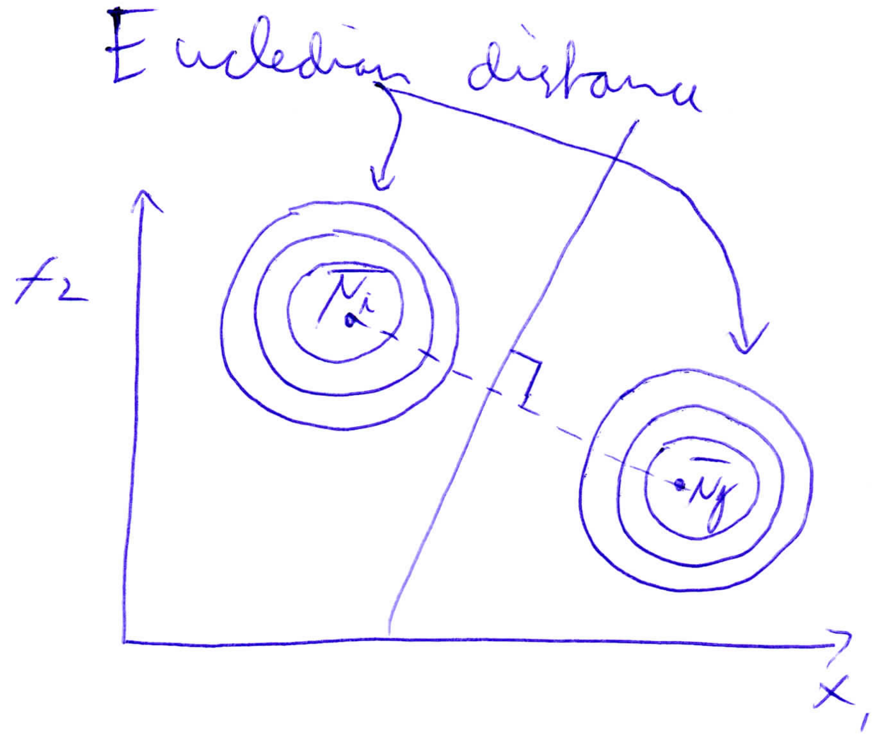
$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \qquad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the Euclidean distance.

This is known as the minimum distance classifier.


Euclidean distance
μ_i
μ_j
x_2
x_1

Case 2: Common covariance matrix

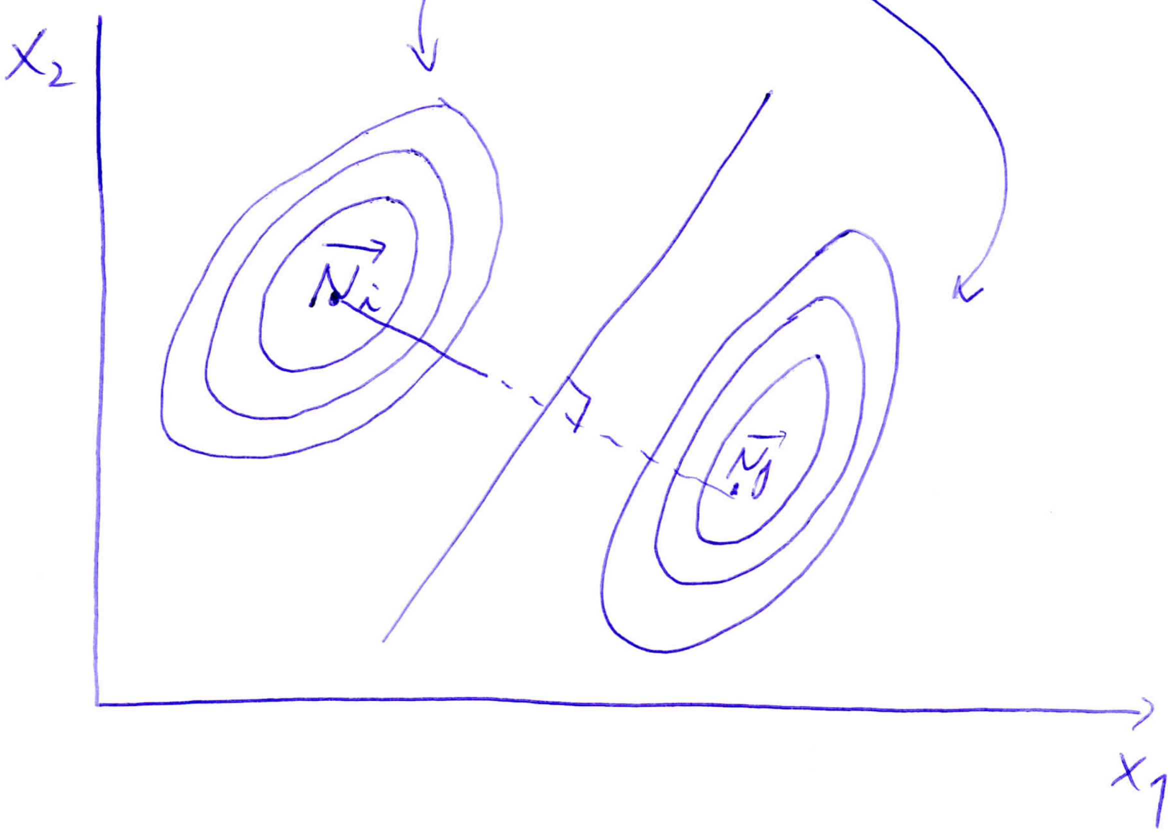
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1,5 \quad 0,3]^T$

$\bar{\mu}_2 = [4,3 \quad 1,3]^T$

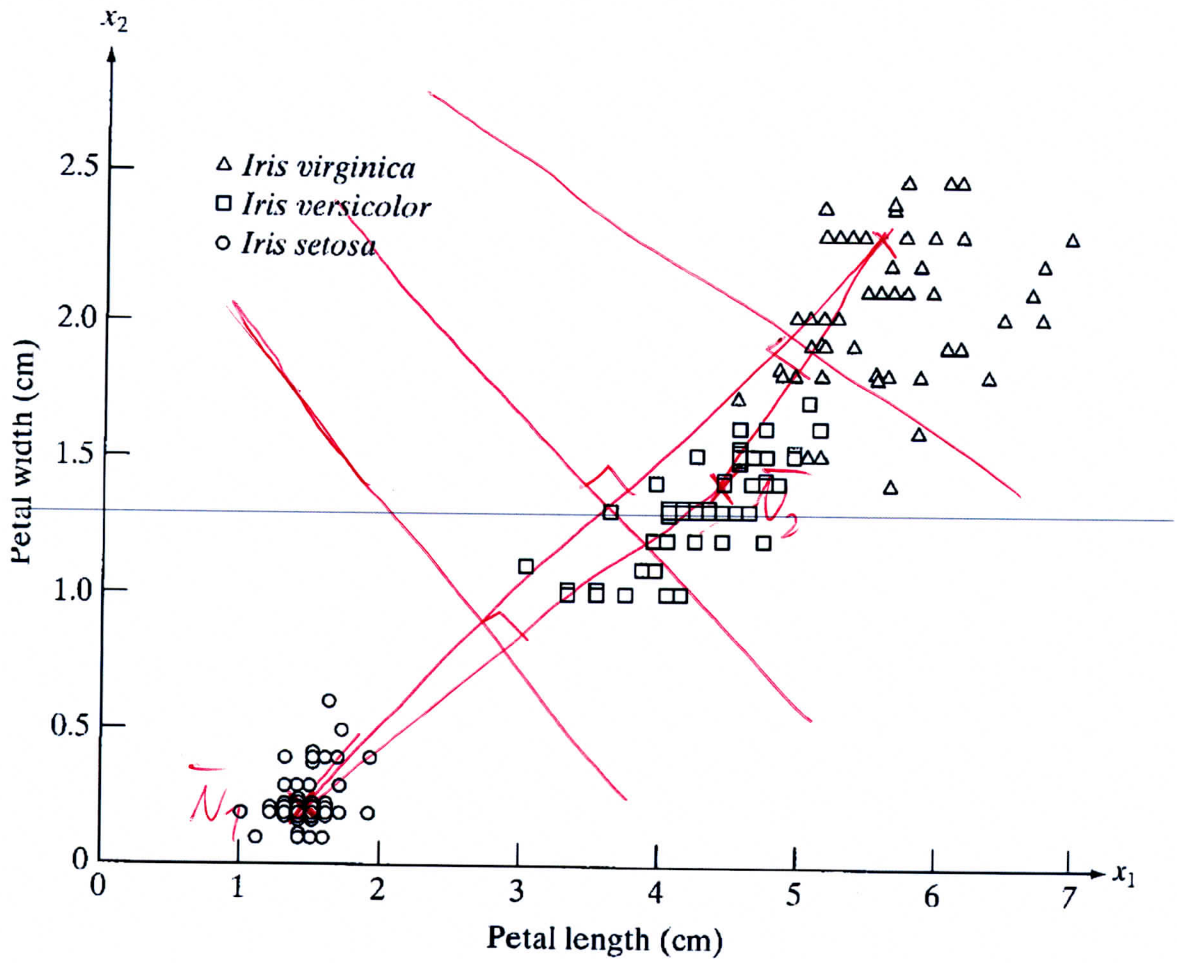
$\bar{\mu}_3 = [5,5 \quad 2,1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$. So we get:

$$G_j^2(x) = \|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2(\bar{x}^T\bar{\mu}_j + \tfrac{1}{2}\mu_j^T\mu_j)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~when c is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2}\begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1,5 & 1,4 \end{bmatrix}\begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2}\begin{bmatrix} 1,07 & 0,54 \end{bmatrix}\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

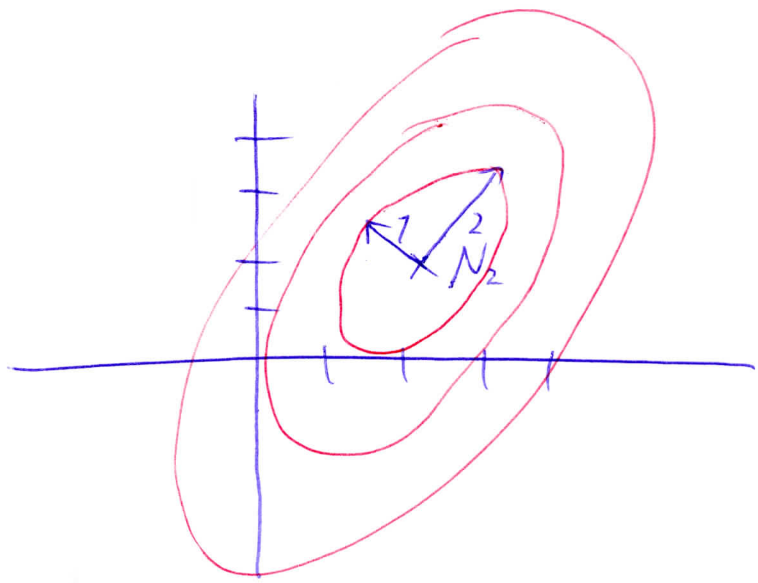
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1 \quad , \quad \lambda_2 = 2$



Ex 5) We have

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

$$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2{,}2 \end{bmatrix} \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$$g_2(\bar{x}) = -1{,}836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x} - \bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

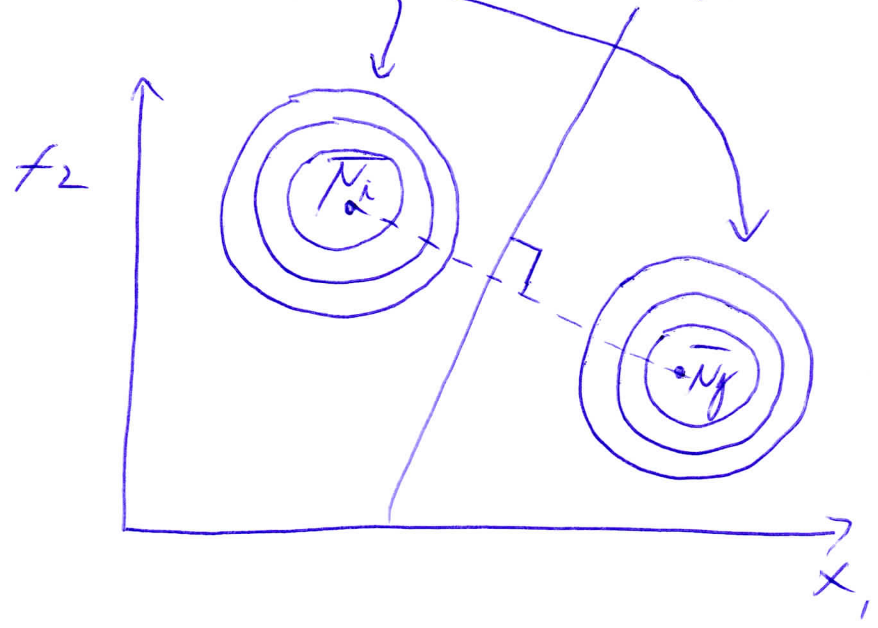
We can discard everything that is common for all classes, so we get:

$$\boxed{g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j)} \qquad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the Euclidean distance.

This is known as the minimum distance classifier.

Eucledian distance



Case 2: Common covariance matrix

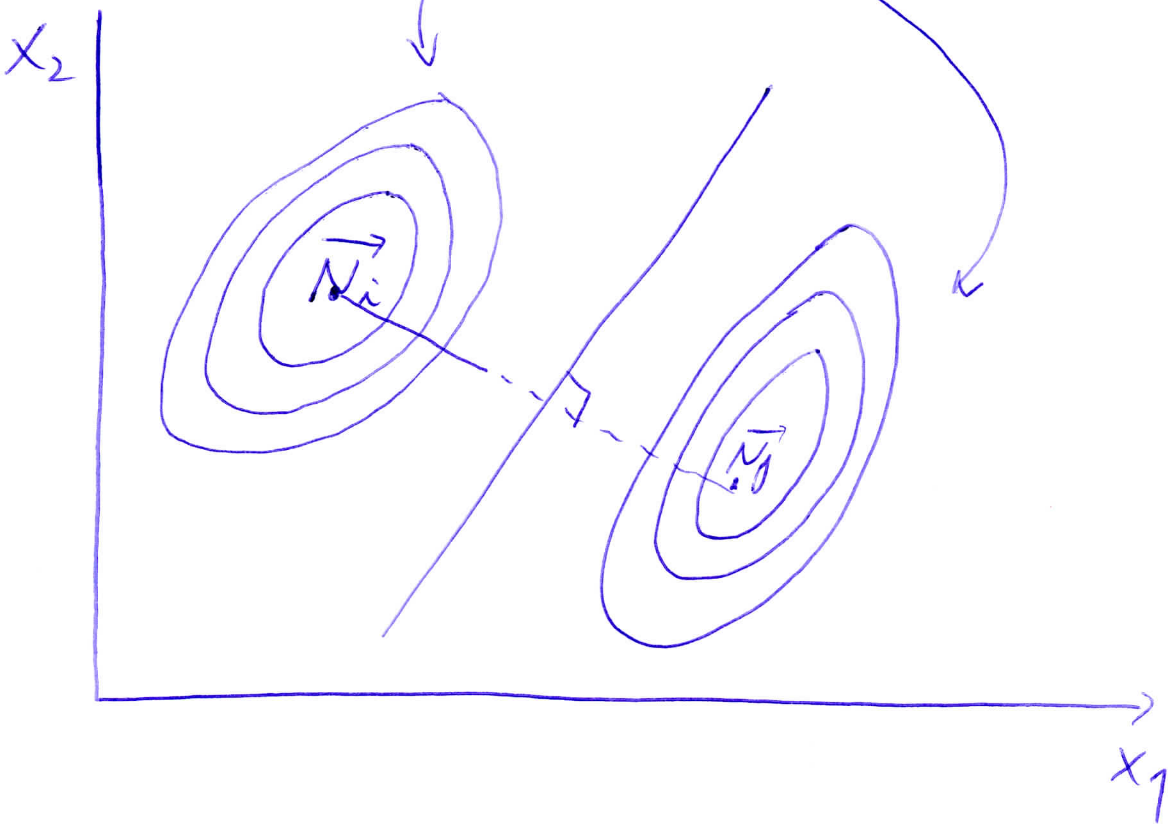
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

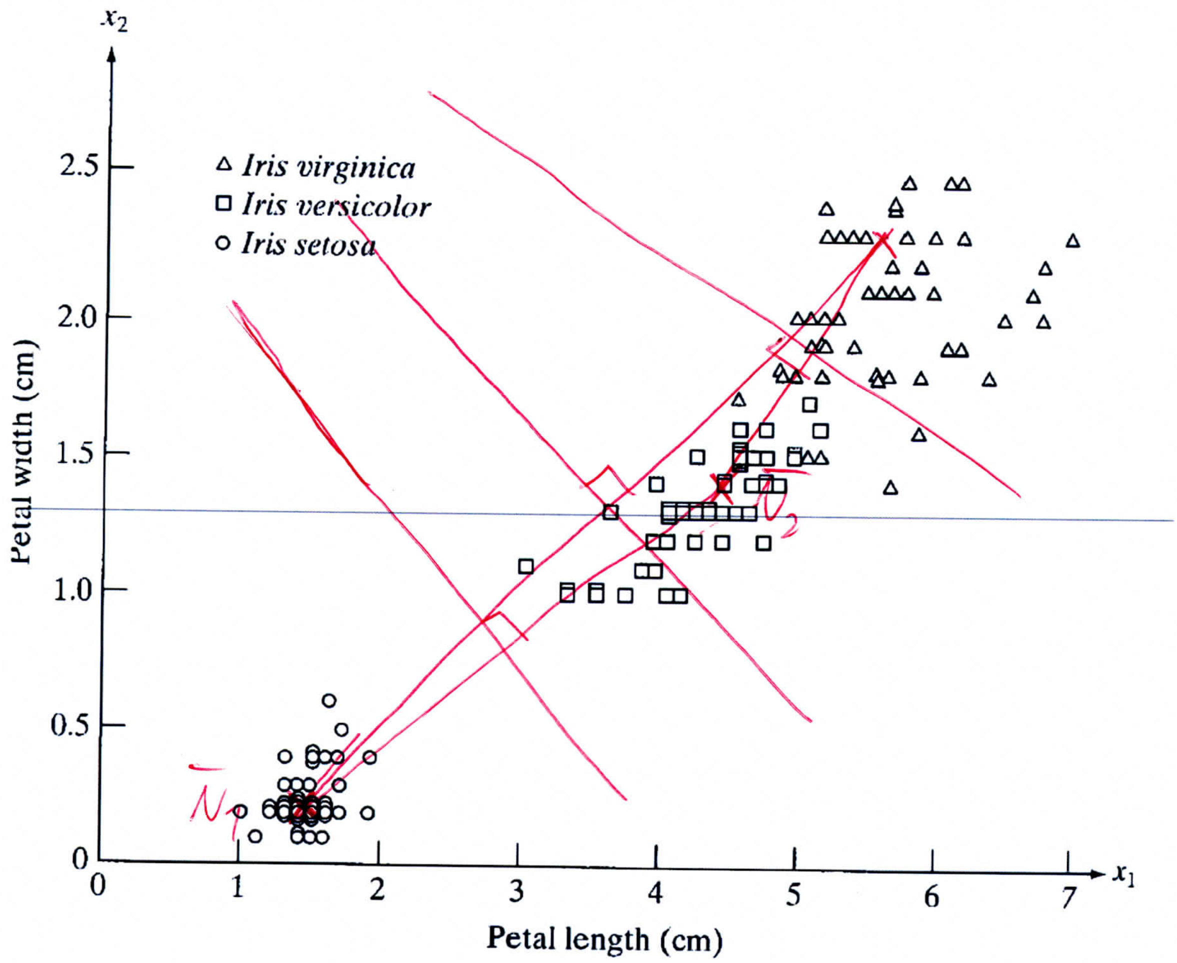
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T (x - \mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \| \bar{x} - \bar{\mu}_j \|^2 = (\bar{x} - \bar{\mu}_j)^T (\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T \bar{x} - 2 \bar{x}^T \bar{\mu}_j + \mu_j^T \mu_j$$

$$= \bar{x}^T \bar{x} - 2 (\bar{x}^T \bar{\mu}_j + \tfrac{1}{2} \mu_j^T \mu_j)$$

$\bar{x}^T \bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j$.

Ex4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where c is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2}\begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2}[1,5 \;\; 1,4]\begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2}[1,07 \;\; 0,54]\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

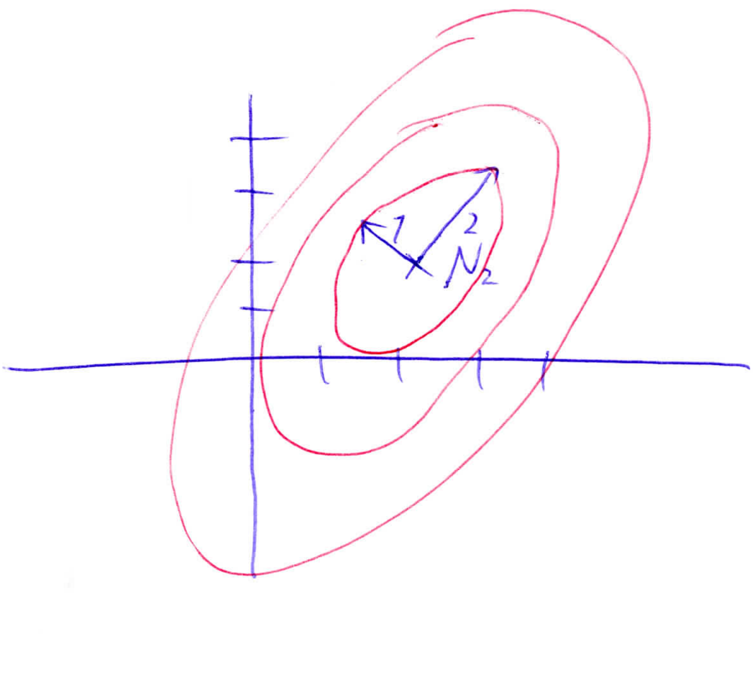
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1$, $\lambda_2 = 2$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2,2 \end{bmatrix} \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0,62 & 1,06 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix}$$

$$= -1,4760$$

$g_2(\bar{x}) = -1,836.$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(w_j|x) = \frac{p(x|w_j)\,P(w_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|w_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|w_j) = \frac{1}{(2\pi)^{n/2}\,|\Sigma_j|^{\frac{1}{2}}}\, e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}. \text{ And}$$

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(w_i|\bar{x}) = \frac{p(\bar{x}|w_j)\,P(w_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x} | \omega_j) P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln (p(\bar{x} | \omega_j) P(\omega_j)) = \ln (p(\bar{x} | \omega_j)) + \ln (P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x} - \bar{\mu}_j) - \frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

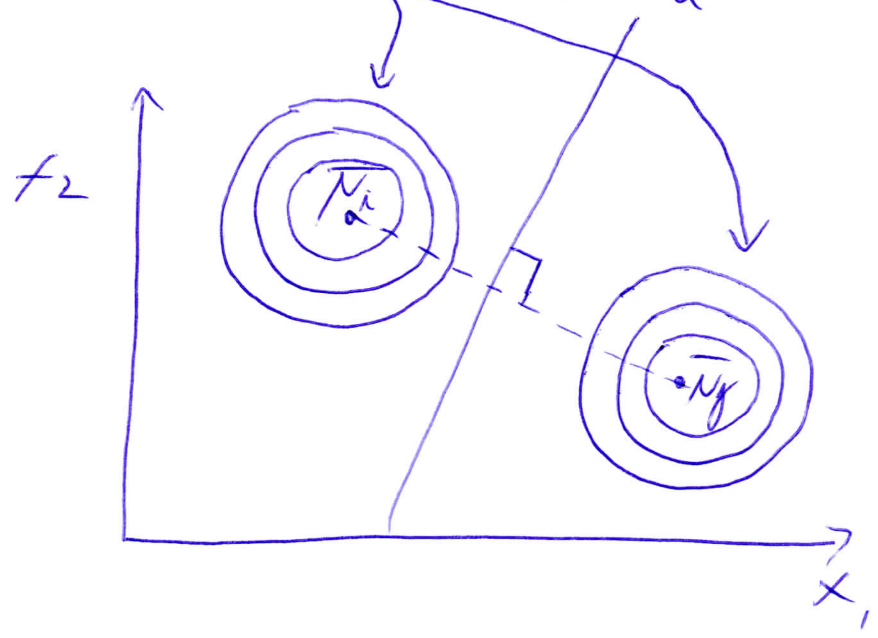
We can discard everything that is common for all classes, so we get:

$$\boxed{g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j)} \qquad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.

Eucledian distance



Case 2: Common covariance matrix

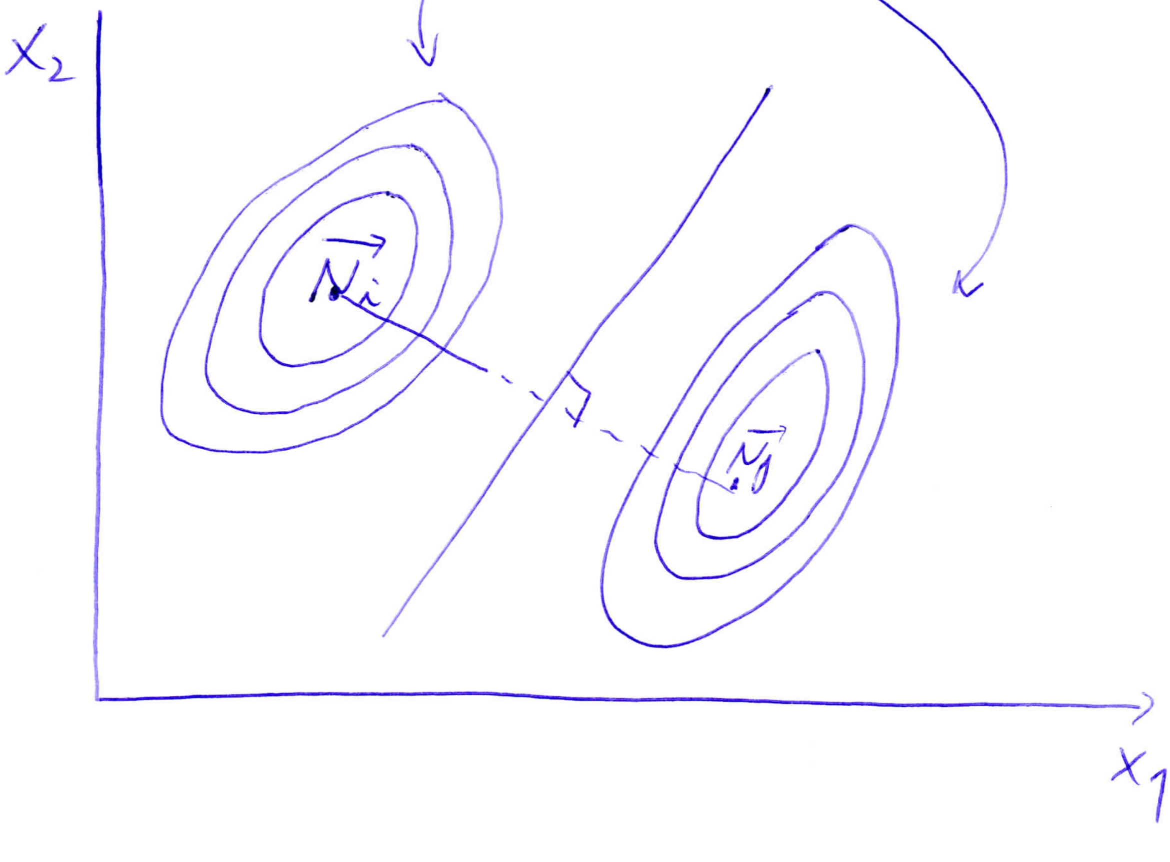
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))}$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \;\; 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \;\; 1{,}3]^T$

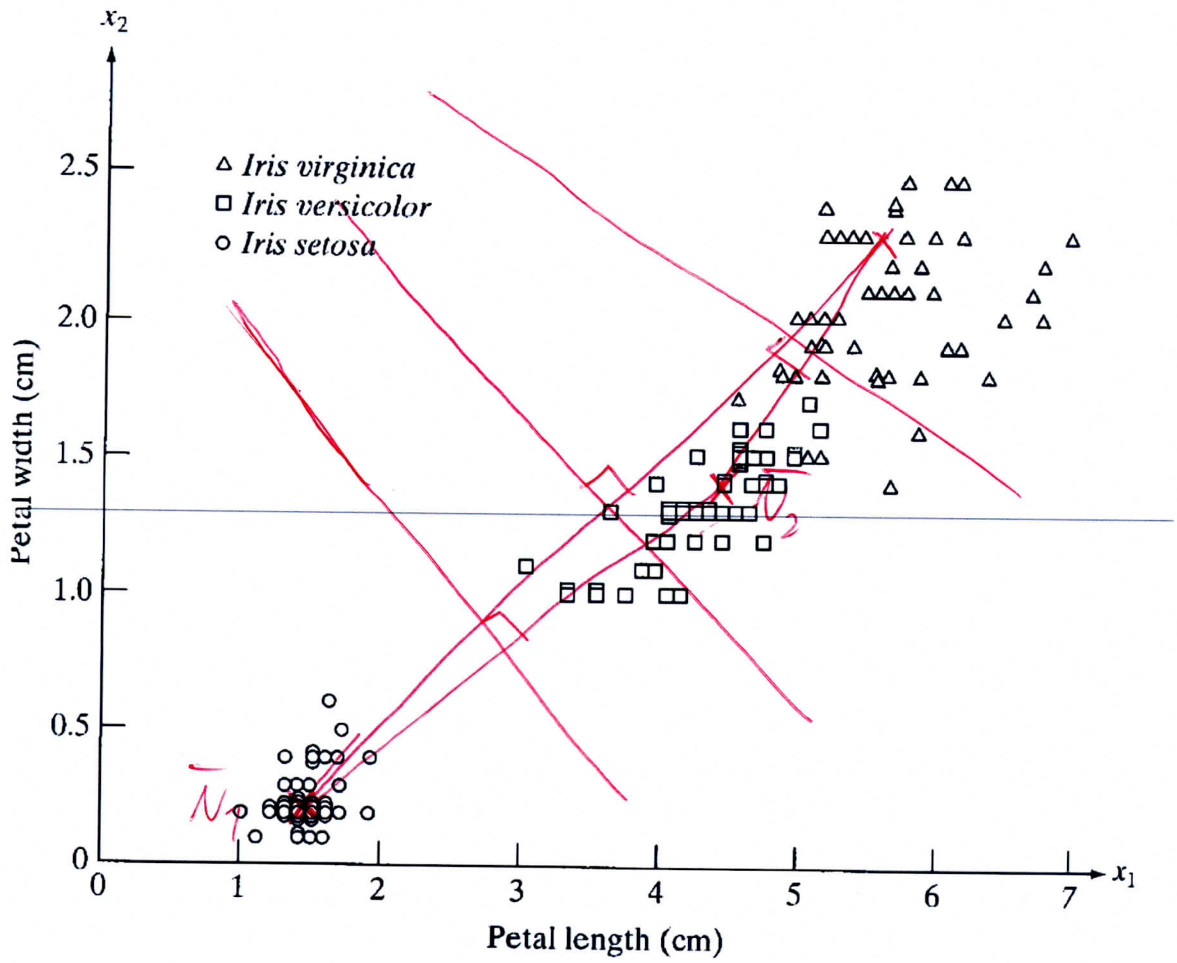
$\bar{\mu}_3 = [5{,}5 \;\; 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G^2_j(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T(\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2(\bar{x}^T\bar{\mu}_j + \frac{1}{2}\mu_j^T\mu_j)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~when C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} \begin{bmatrix} 1,5 & 1,4 \end{bmatrix} \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2} \begin{bmatrix} 1,07 & 0,54 \end{bmatrix} \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

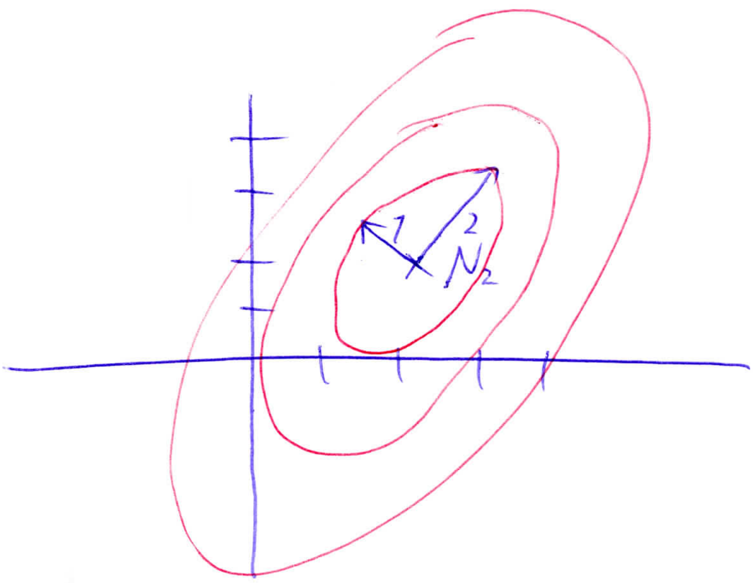
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1$, $\lambda_2 = 2$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2,2 \end{bmatrix}\begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}\begin{bmatrix} 1 \\ 2,2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0,62 & 1,06 \end{bmatrix}\begin{bmatrix} 1 \\ 2,2 \end{bmatrix}$$

$$= -1,4760$$

$g_2(\bar{x}) = -1,836.$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \quad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}. \text{ And}$$

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

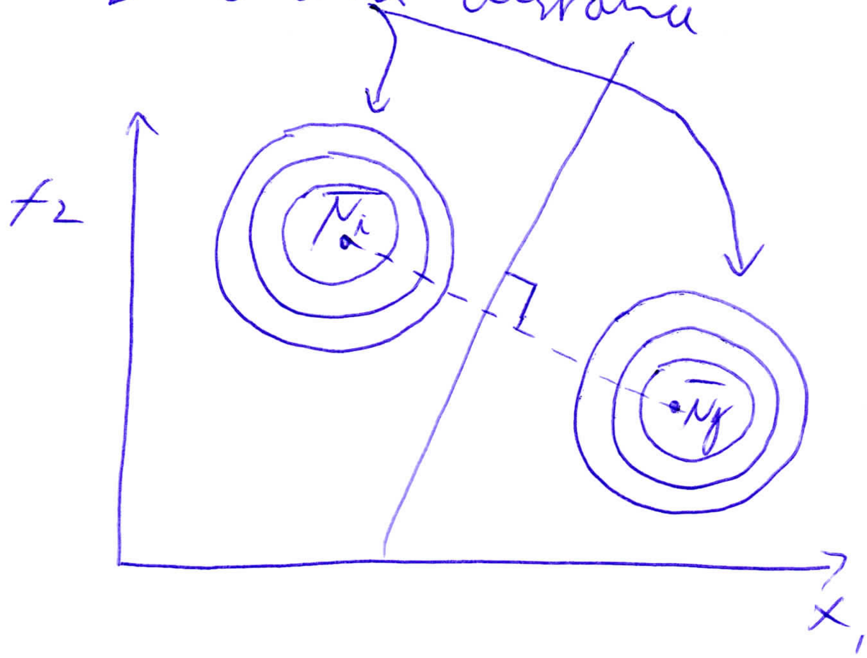
We can discard everything that is common for all classes, so we get:

$$\boxed{g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j)} \quad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.

Eucledian distance



Case 2: Common covariance matrix

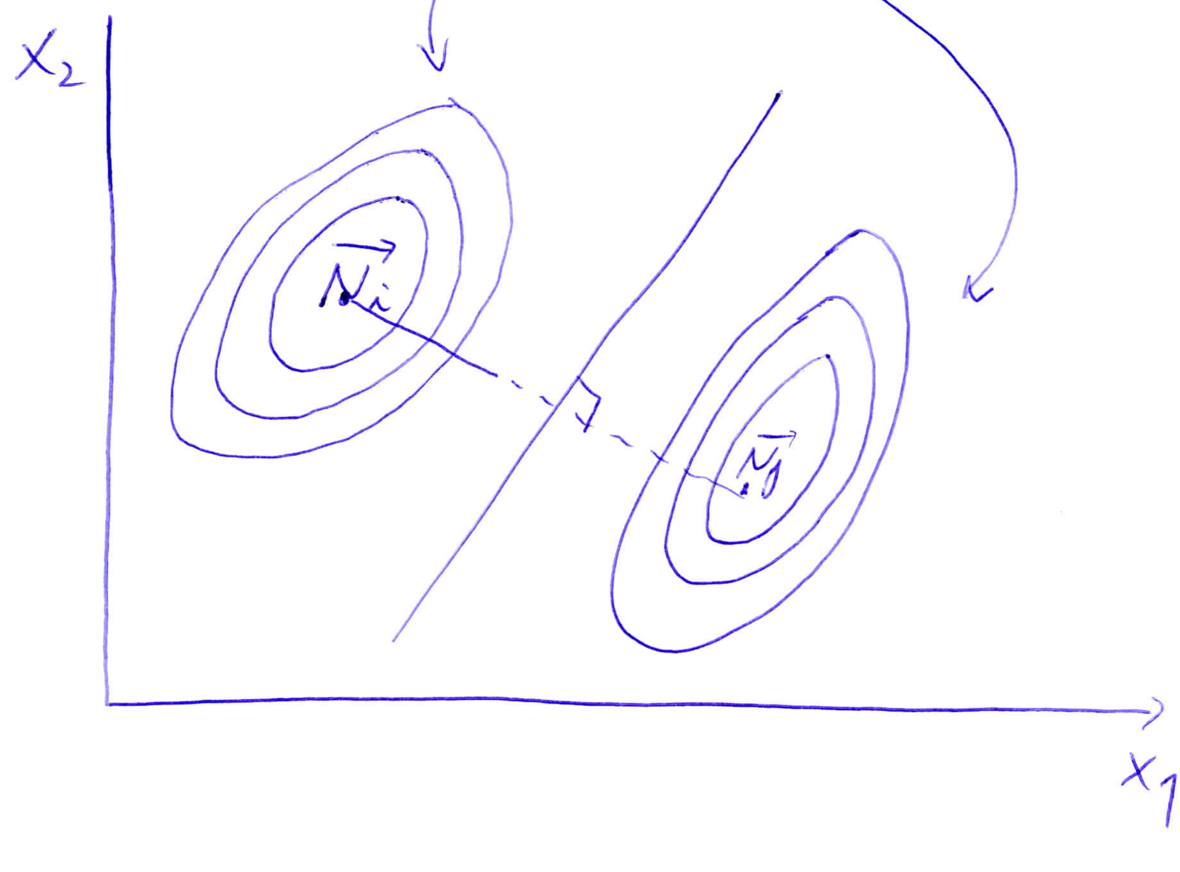
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

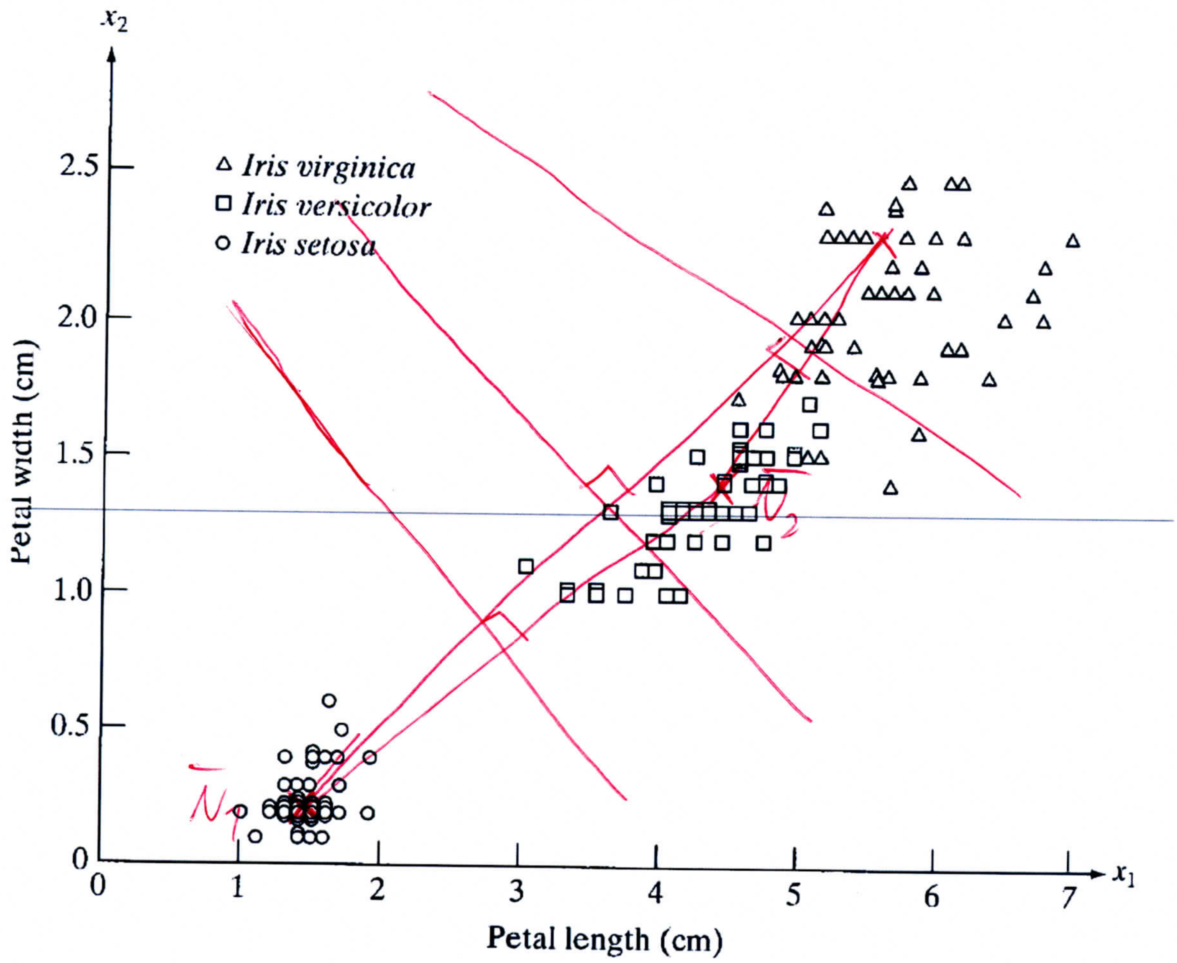
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T(\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2(\bar{x}^T\bar{\mu}_j + \tfrac{1}{2}\mu_j^T\mu_j)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2}\begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \ \ 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2}[1,5 \ \ 1,4]\begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2}[1,07 \ \ 0,54]\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0{,}1205$

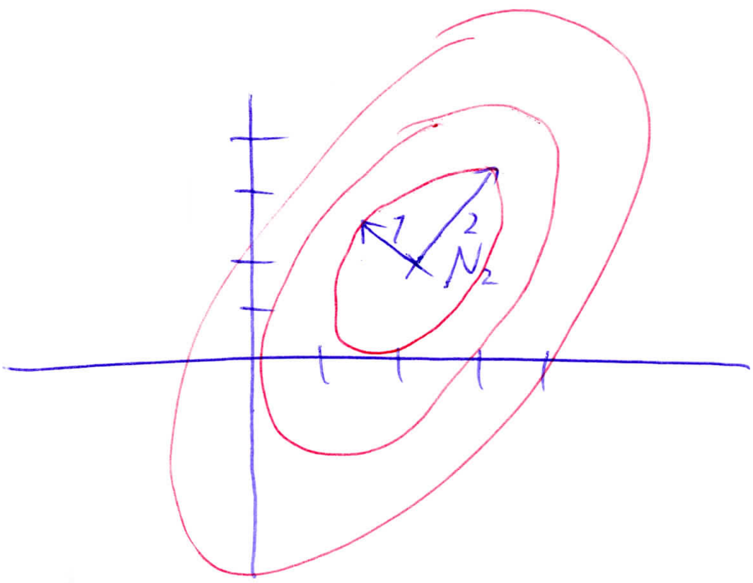
$g_3(\bar{x}) = -4{,}7095$

$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b) $\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$

Eigenvectors: $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1$, $\lambda_2 = 2$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2{,}2 \end{bmatrix}\begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix}\begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$g_2(\bar{x}) = -1{,}836.$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}. \text{ And}$$

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

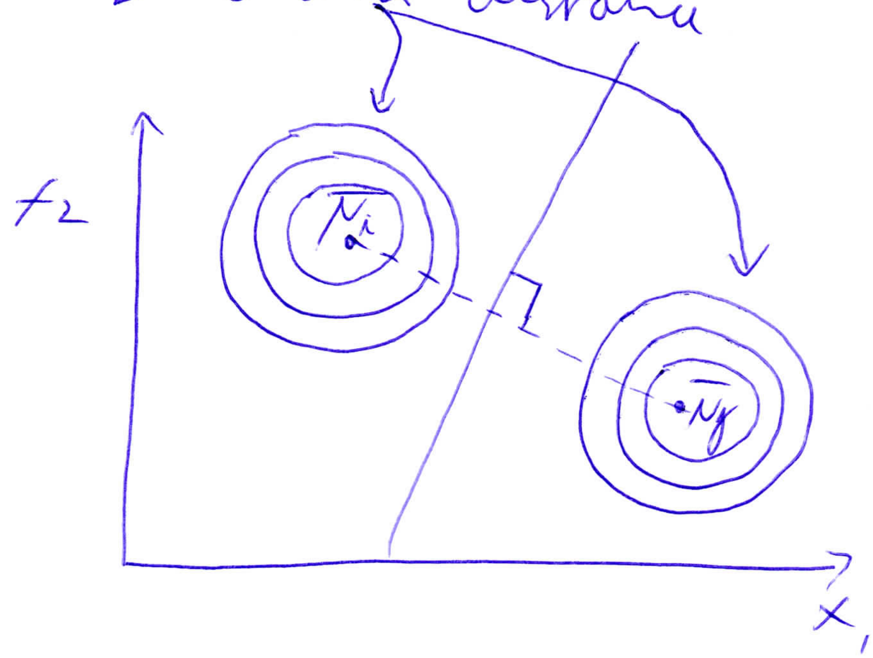
We can discard everything that is common for all classes, so we get:

$$\boxed{g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j)} \quad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.

Euclidian distance



Case 2: Common covariance matrix

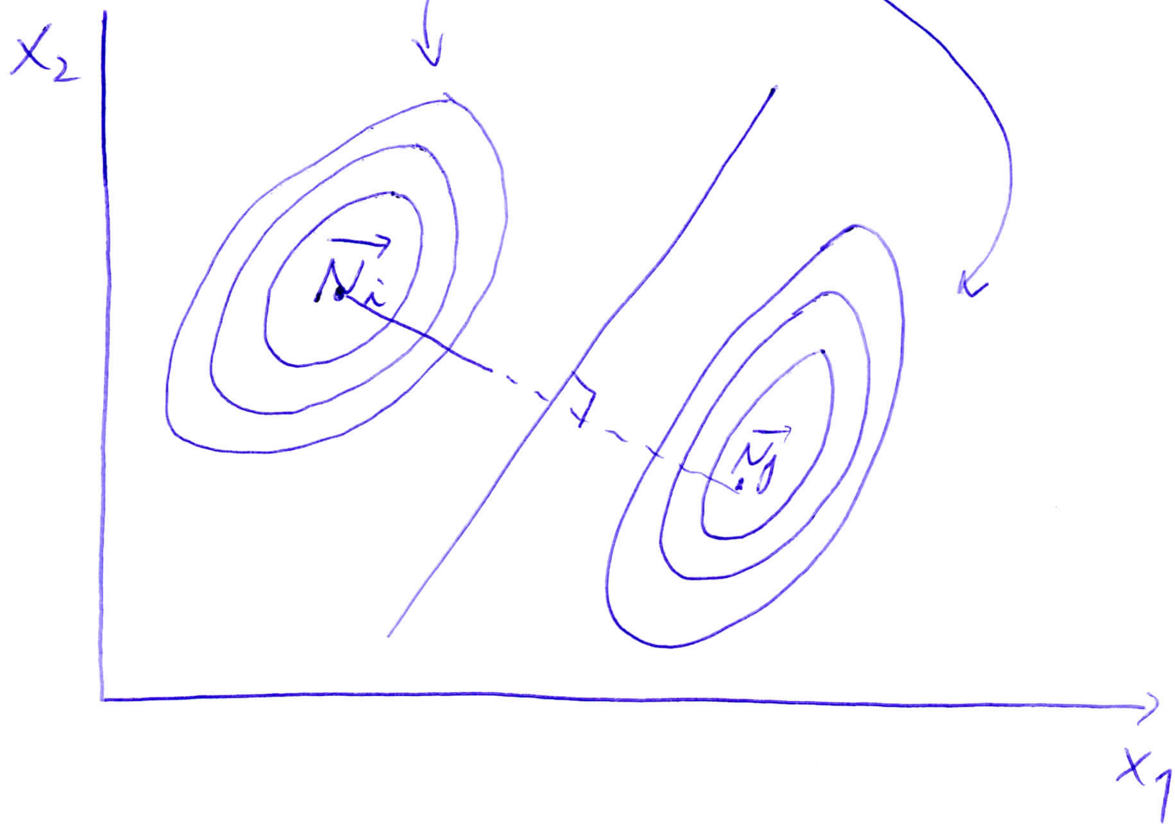
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

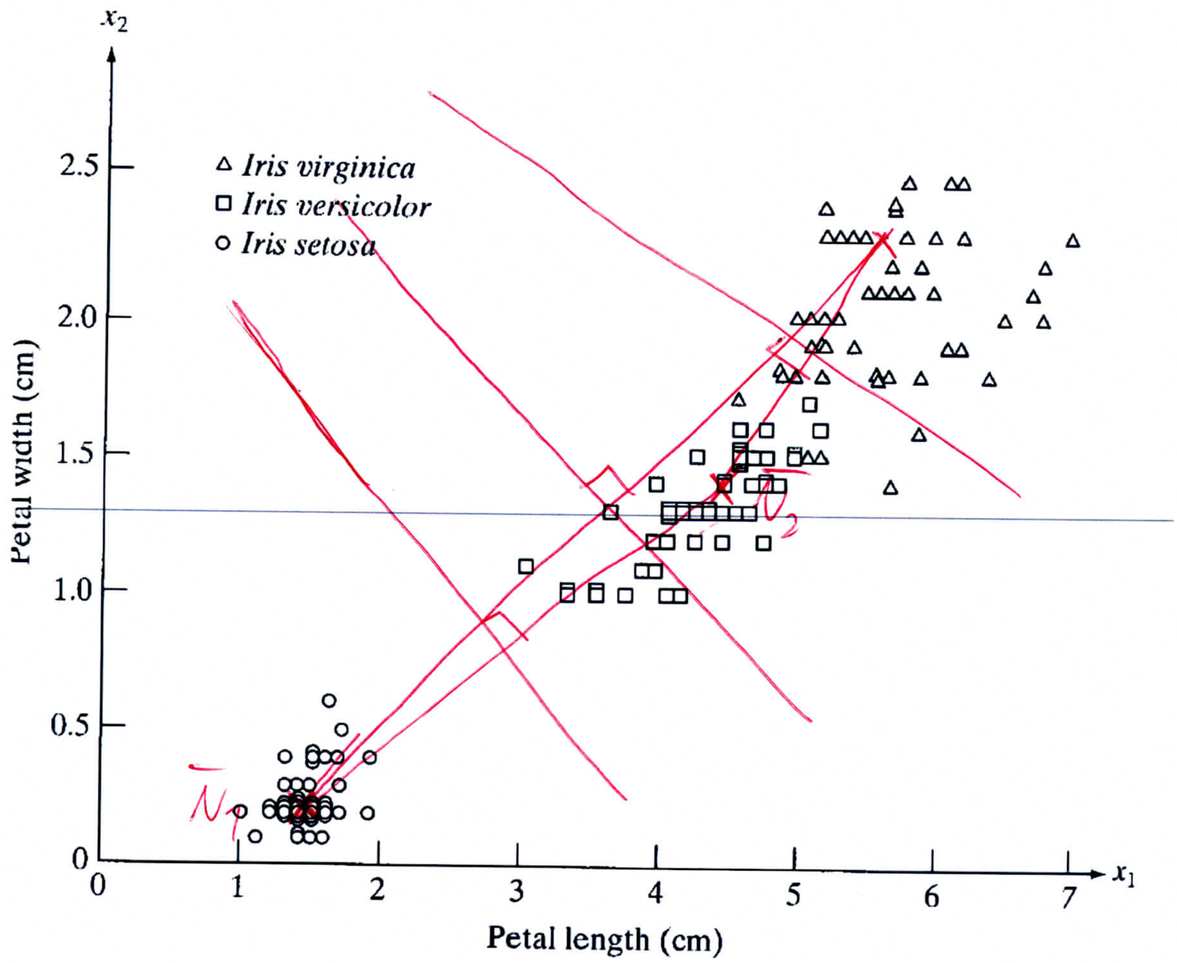
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T (x-\mu_j)}$

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

So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T (\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T \bar{x} - 2\bar{x}^T \bar{\mu}_j + \mu_j^T \mu_j$$

$$= \bar{x}^T \bar{x} - 2\left(\bar{x}^T \bar{\mu}_j + \tfrac{1}{2} \mu_j^T \mu_j\right)$$

$\bar{x}^T \bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2{,}1 \\ 1{,}9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1{,}5 \\ 2{,}0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1{,}8 & -0{,}4 \\ -0{,}4 & 1{,}2 \end{bmatrix} = \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix}.$$

Given $\bar{x} = [1{,}6 \;\; 1{,}5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} [1{,}5 \;\; 1{,}4] \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix}$$

$$= -\frac{1}{2} [1{,}07 \;\; 0{,}54] \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix} = \underline{-1{,}1805}$$

$g_2(\bar{x}) = -0,1205$

$g_3(\bar{x}) = -4,7095$

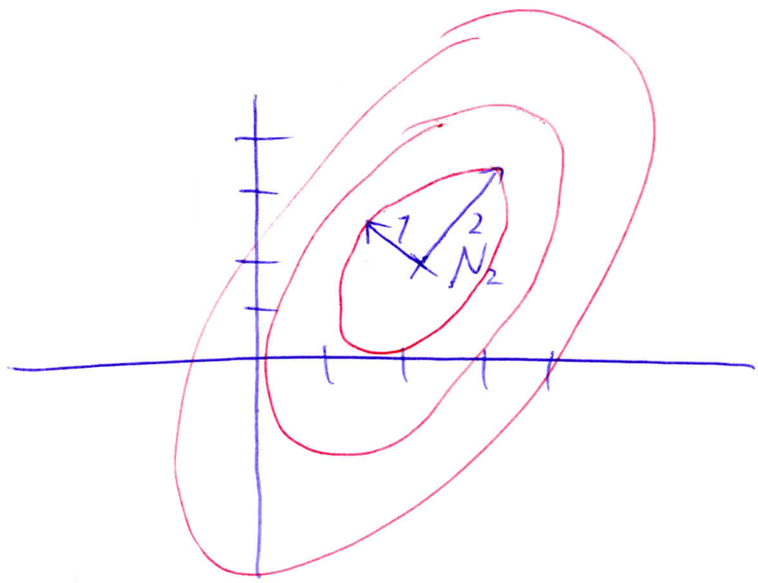
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}$$

Eigenvectors : $v_1 = \begin{bmatrix} -0,895 \\ 0,447 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0,447 \\ 0,895 \end{bmatrix}$

$\lambda_1 = 1$, $\lambda_2 = 2$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}$, $\Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix} \left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2} \begin{bmatrix} 1 & 2{,}2 \end{bmatrix} \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$g_2(\bar{x}) = -1{,}836.$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \quad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}. \text{ And}$$

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

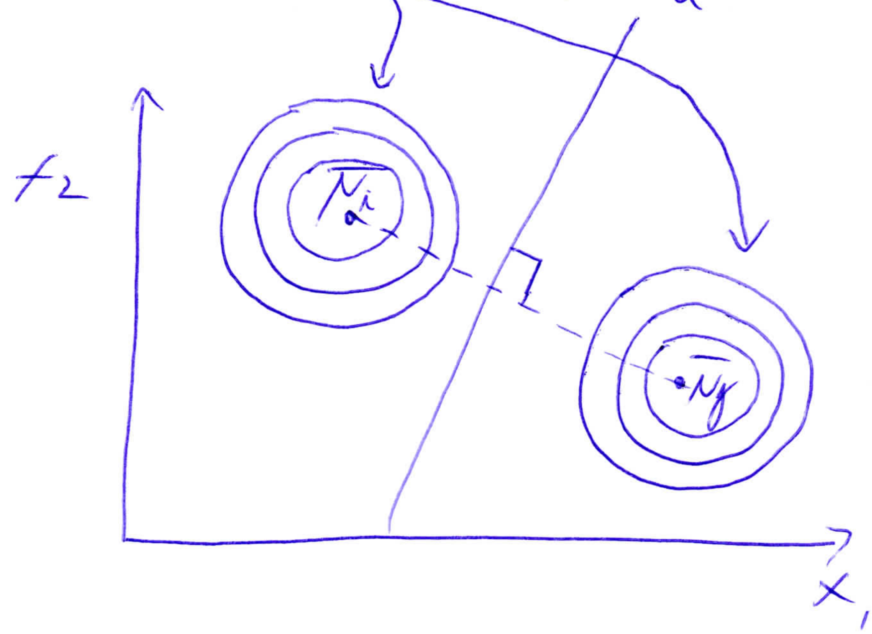
We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \quad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the Euclidean distance.

This is known as the minimum distance classifier.

Euclidean distance



Case 2: Common covariance matrix

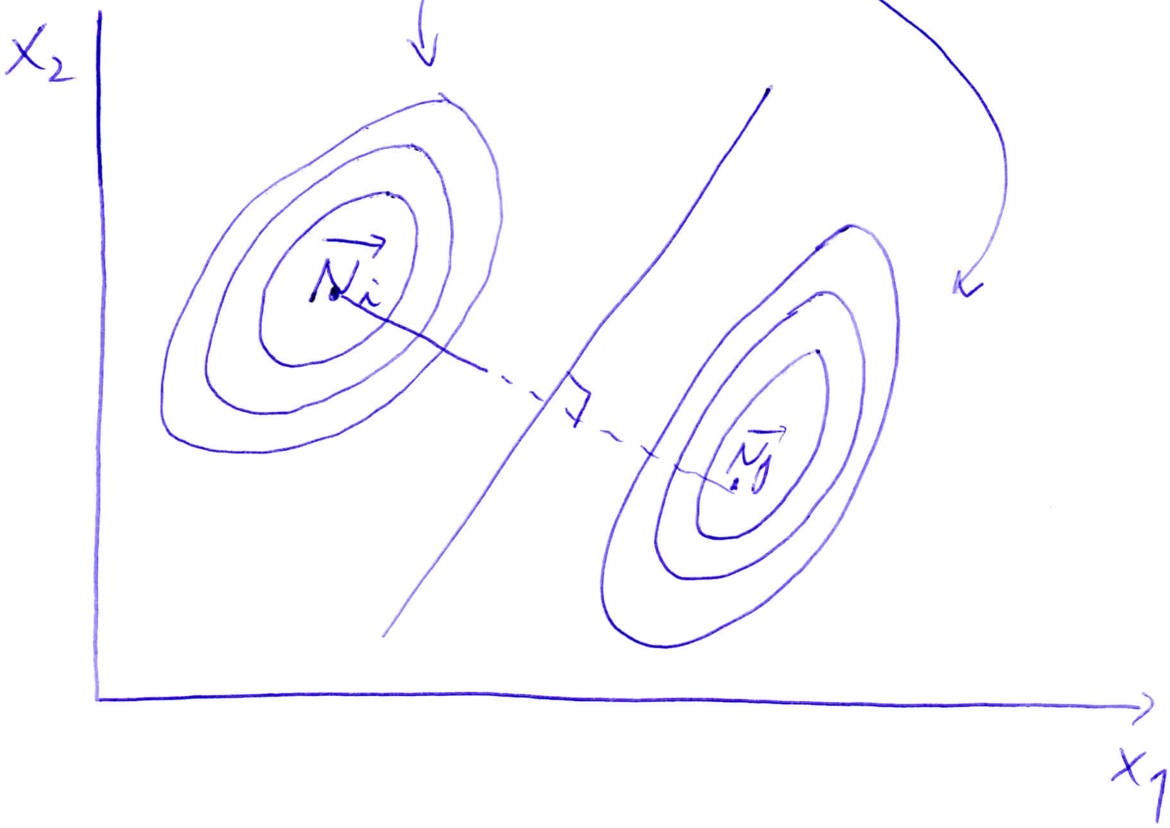
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

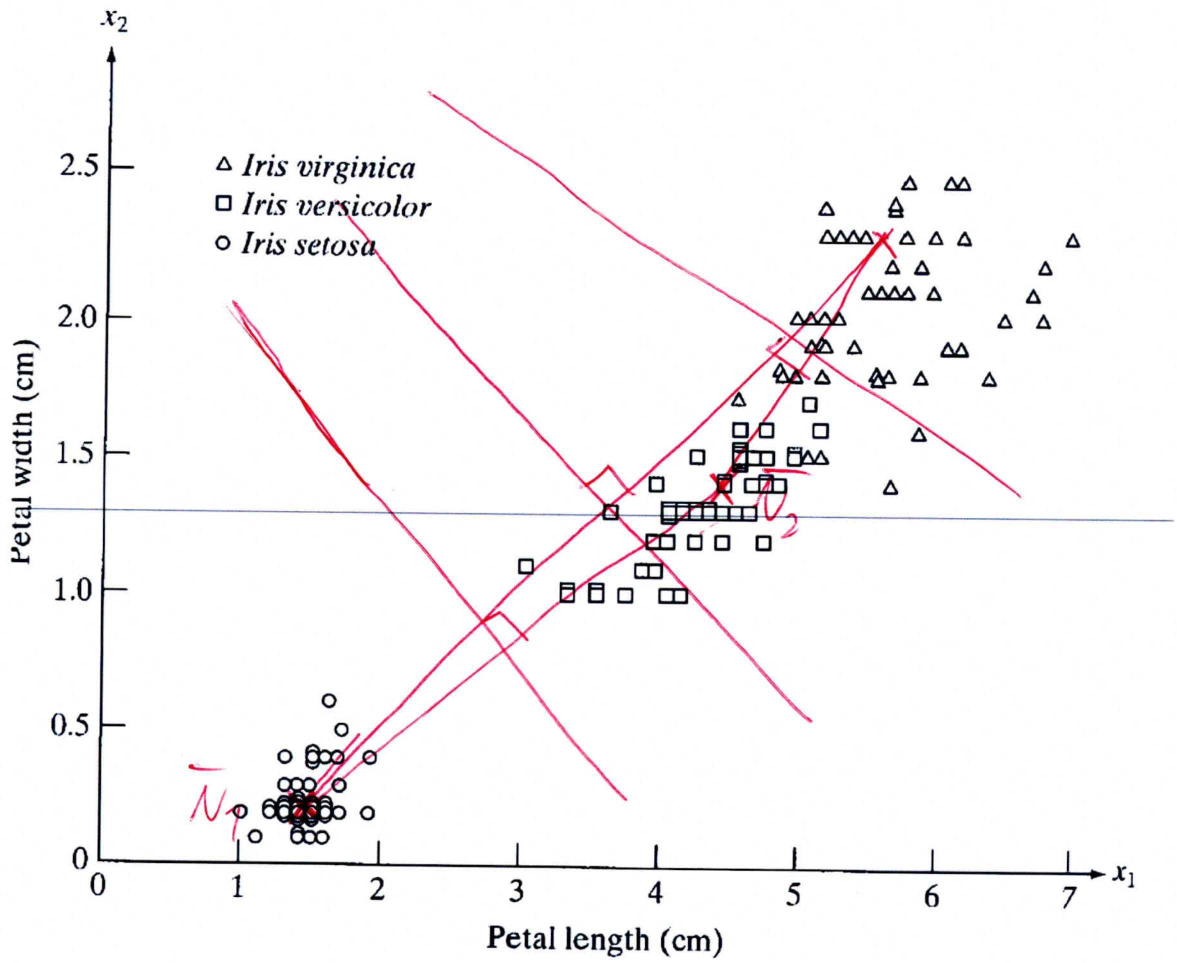
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G^2_j(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T(\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2(\bar{x}^T\bar{\mu}_j + \tfrac{1}{2}\mu_j^T\mu_j)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where c is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} [1,5 \;\; 1,4] \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2} [1,07 \;\; 0,54] \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0,1205$

$g_3(\bar{x}) = -4,7095$

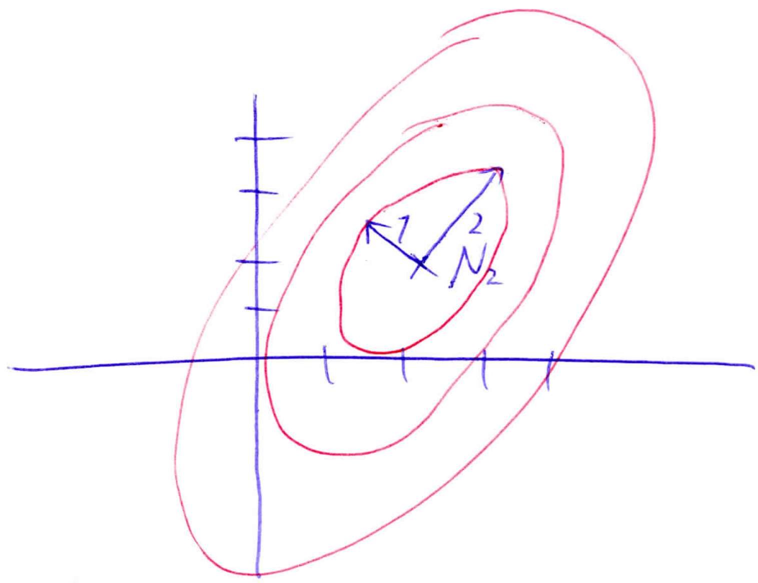
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}$$

Eigenvectors : $v_1 = \begin{bmatrix} -0,895 \\ 0,447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0,447 \\ 0,895 \end{bmatrix}$

$\lambda_1 = 1 \quad , \quad \lambda_2 = 2$



Ex 5) We have

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

$$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2} \begin{bmatrix} 1 & 2,2 \end{bmatrix} \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} 0,62 & 1,06 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix}$$

$$= -1,4760$$

$$g_2(\bar{x}) = -1,836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j | x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \quad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2} |\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N\cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

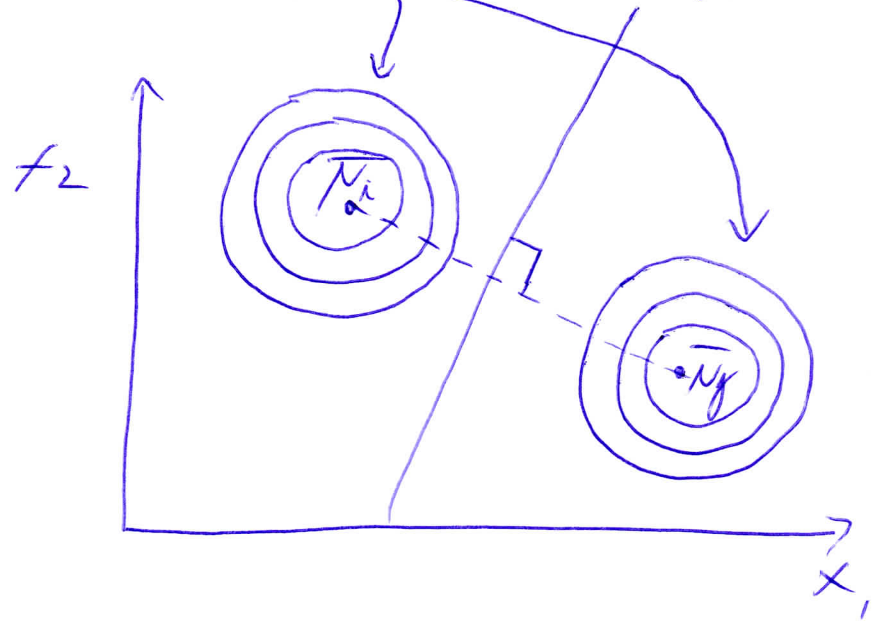
We can discard everything that is common for all classes, so we get:

$$\boxed{g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j)} \quad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.

Eucledian distance



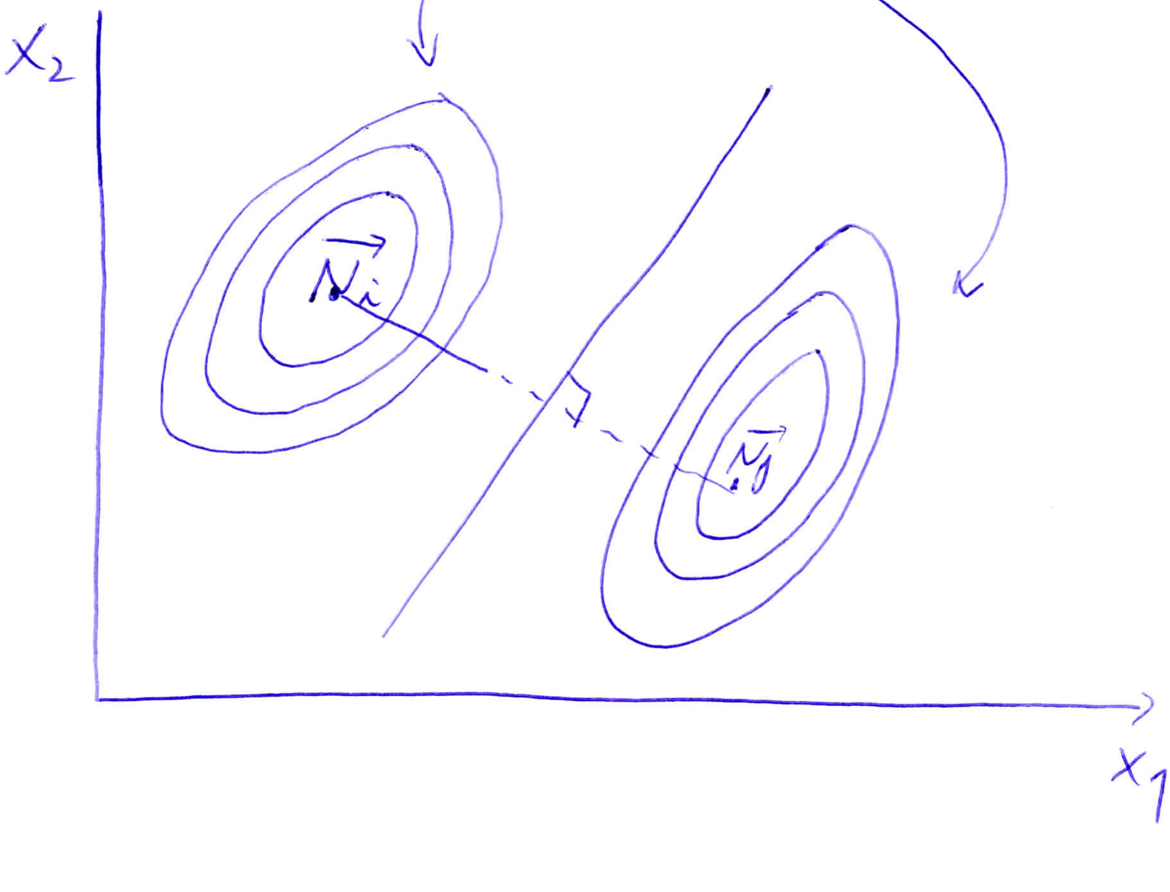
Case 2: Common covariance matrix

Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$\boxed{\begin{array}{l} g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j)) \\[1em] (\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) \text{ is the} \\ \text{Mahalanobis distance between } \bar{x} \text{ and } \bar{\mu} \end{array}}$$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

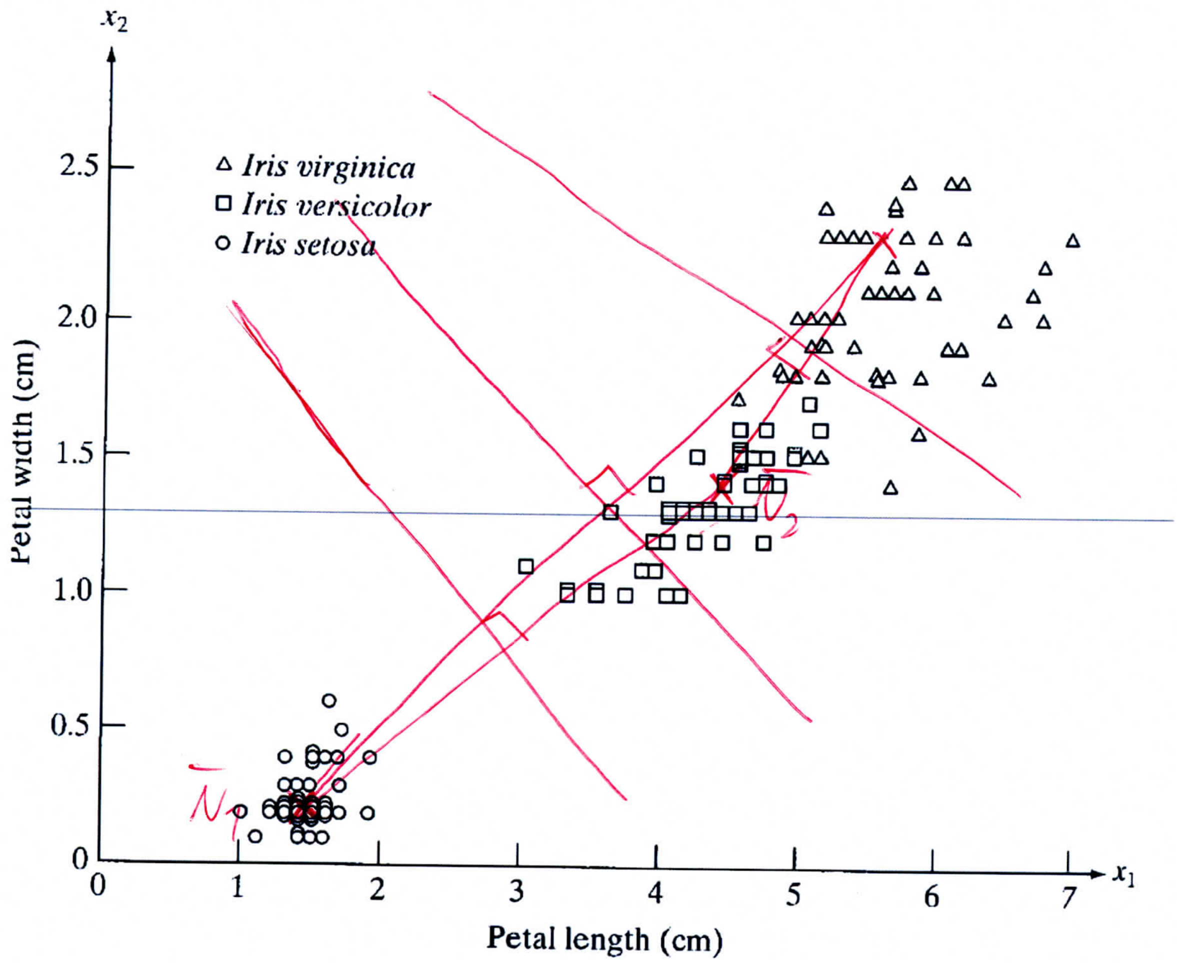
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T(\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2\left(\bar{x}^T\bar{\mu}_j + \frac{1}{2}\mu_j^T\mu_j\right)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2{,}1 \\ 1{,}9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1{,}5 \\ 2{,}0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1{,}8 & -0{,}4 \\ -0{,}4 & 1{,}2 \end{bmatrix} = \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix}.$$

Given $\bar{x} = [1{,}6 \ \ 1{,}5]^T$

$$g_1(\bar{x}) = -\frac{1}{2} \left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix} \right)^T \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix} \right)$$

$$= -\frac{1}{2} \begin{bmatrix} 1{,}5 & 1{,}4 \end{bmatrix} \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix}$$

$$= -\frac{1}{2} \begin{bmatrix} 1{,}07 & 0{,}54 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix} = \underline{-1{,}1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

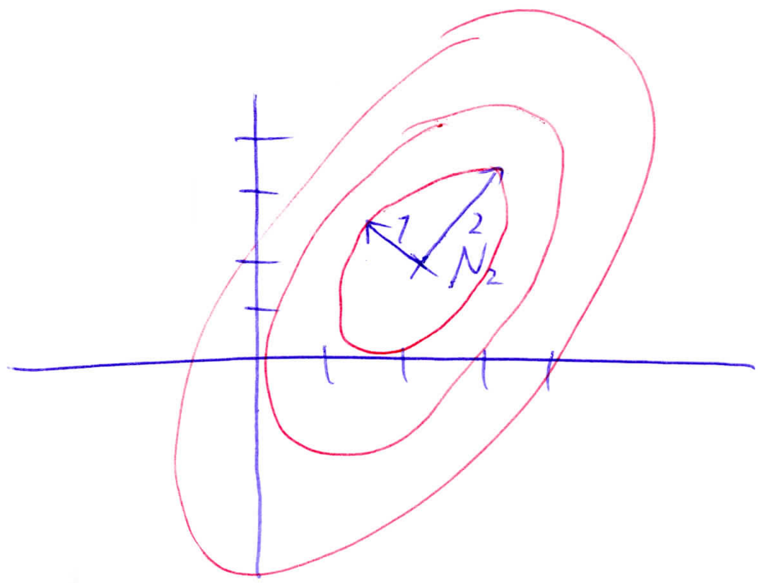
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors : $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1$, $\lambda_2 = 2$



Ex 5) We have

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

$$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2{,}2 \end{bmatrix} \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$$g_2(\bar{x}) = -1{,}836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \quad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

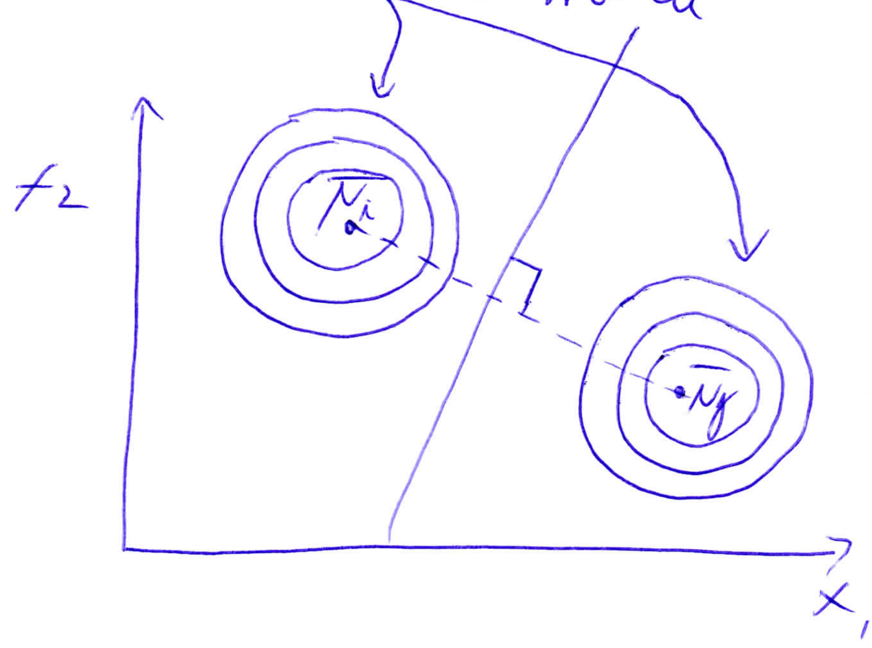
We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \qquad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the Euclidean distance.

This is known as the minimum distance classifier.

Eucledian distance



Case 2: Common covariance matrix

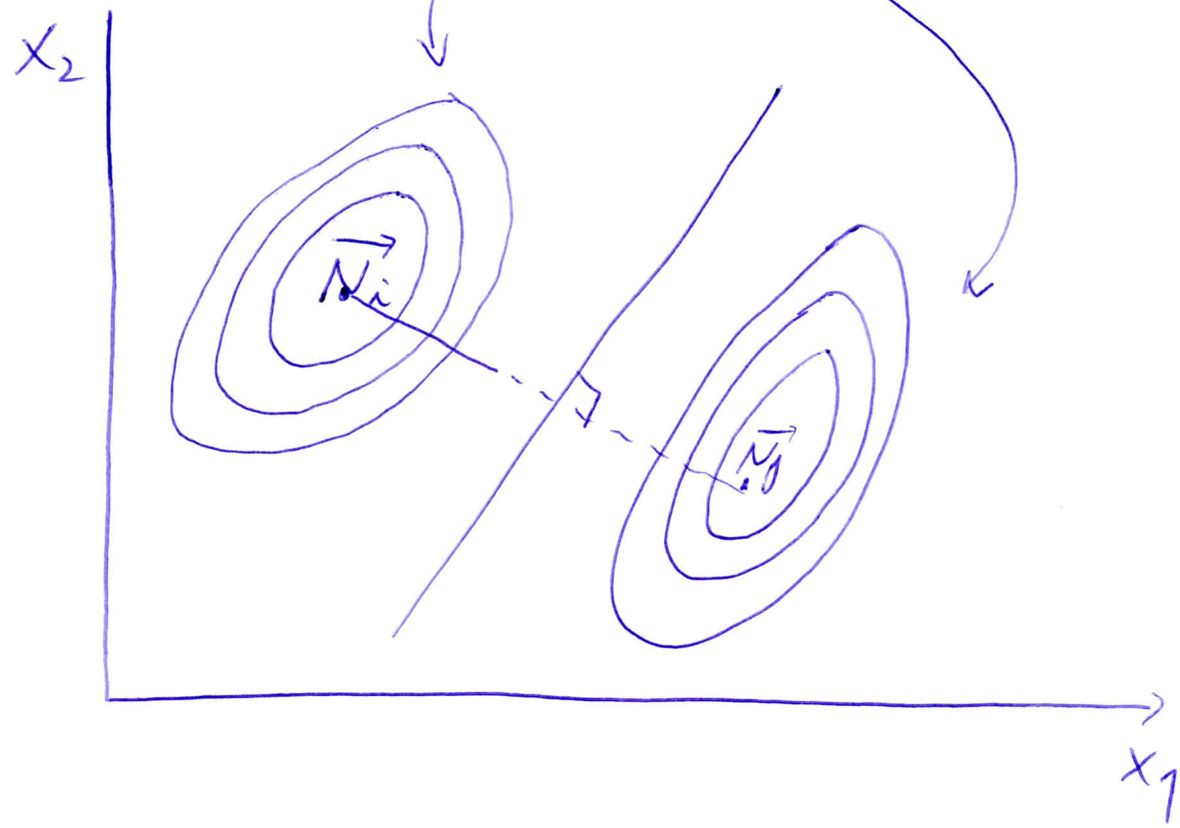
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))}$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

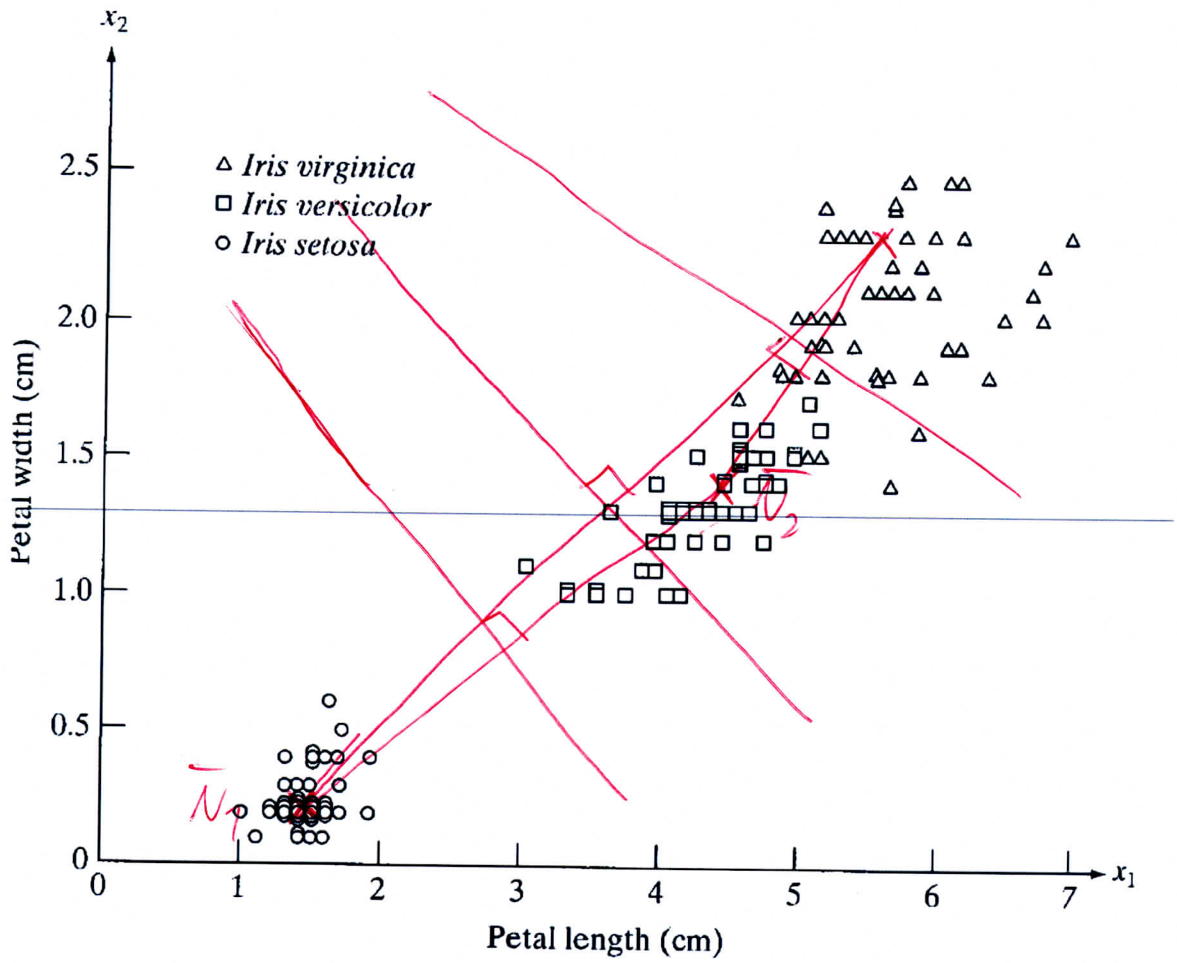
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \|x - \bar{\mu}_j\|^2 = (x - \bar{\mu}_j)^T(\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2\left(\bar{x}^T\bar{\mu}_j + \frac{1}{2}\mu_j^T\mu_j\right)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$, ~~where c is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} [1,5 \;\; 1,4] \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2} [1,07 \;\; 0,54] \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

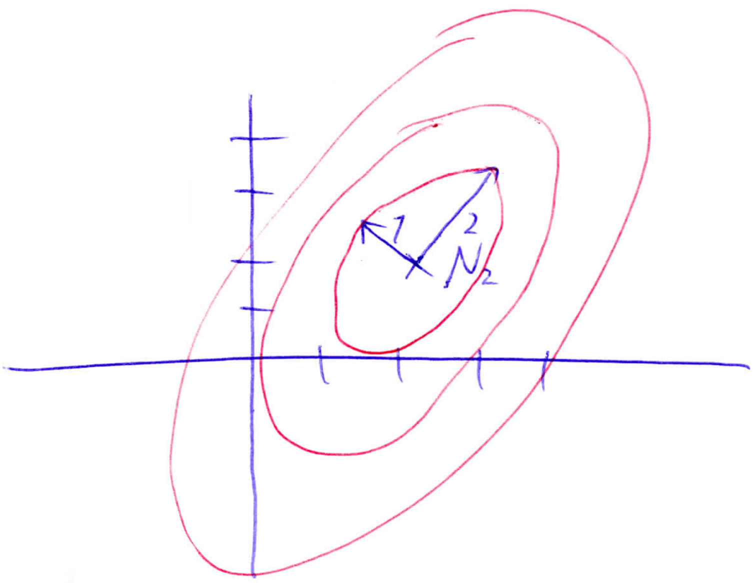
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors : $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1$, $\lambda_2 = 2$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}$, $\Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}\left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2,2 \end{bmatrix}\begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}\begin{bmatrix} 1 \\ 2,2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0,62 & 1,06 \end{bmatrix}\begin{bmatrix} 1 \\ 2,2 \end{bmatrix}$$

$$= -1,4760$$

$g_2(\bar{x}) = -1,836.$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \quad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

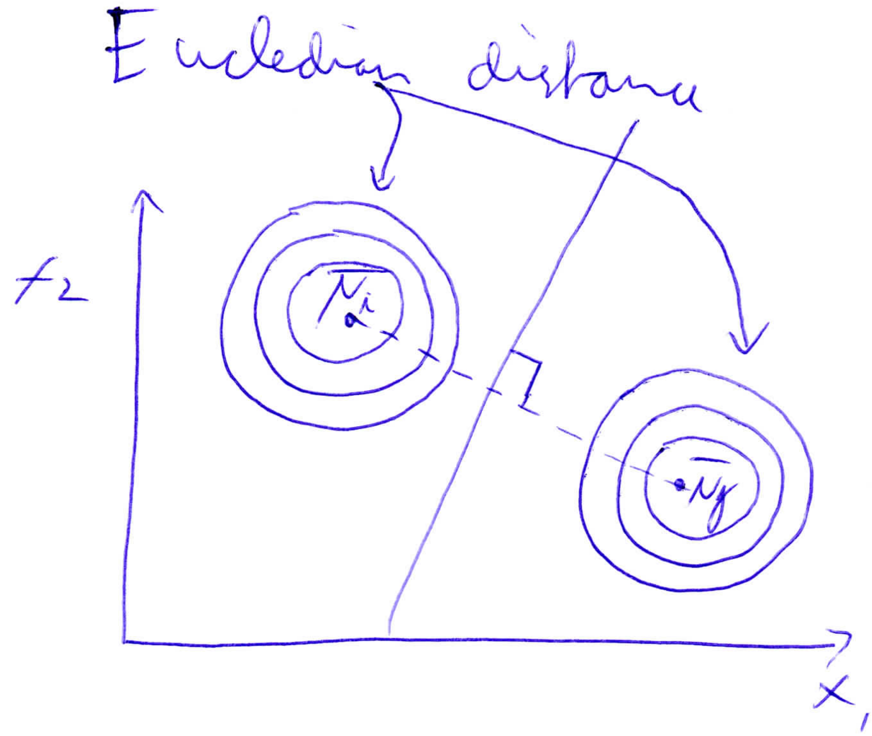
$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \quad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.


Euclidean distance
x_2
μ_i
μ_j
x_1

Case 2: Common covariance matrix

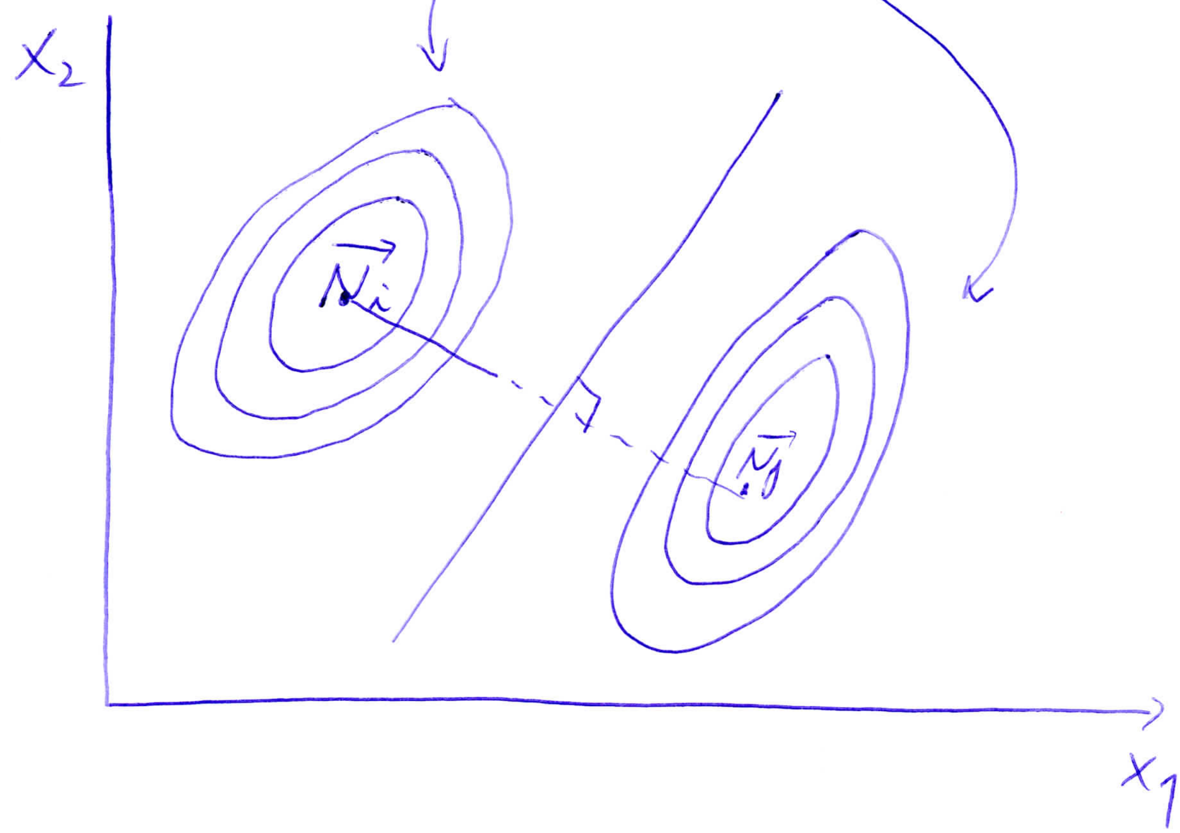
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

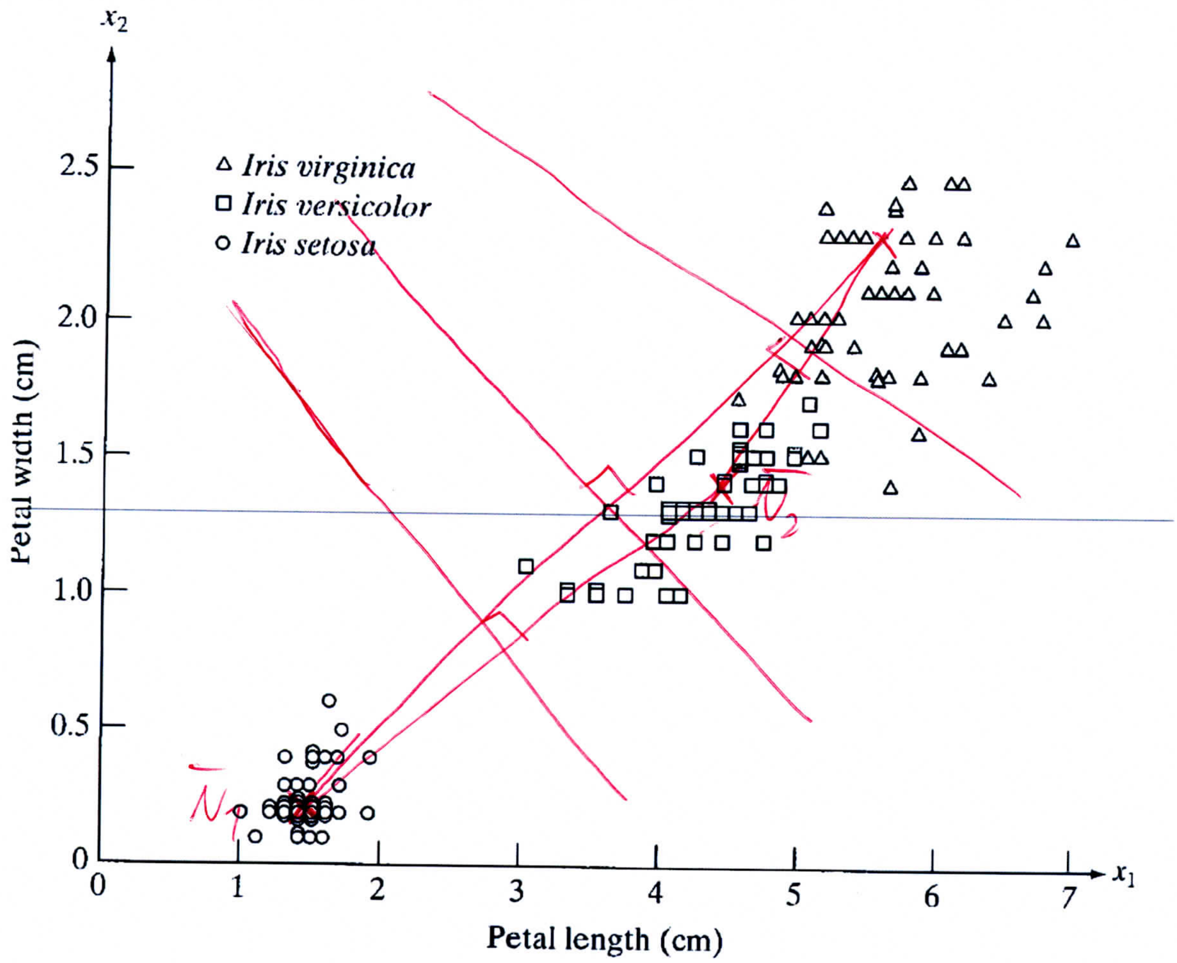
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = ||\bar{x} - \bar{\mu}_j||.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j.$$

Def: $G_j(x) = ||\bar{x} - \bar{\mu}_j|| = \sqrt{(x-\mu_j)^T (x-\mu_j)}$

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

So we get:

$$G_j^2(x) = ||\bar{x} - \bar{\mu}_j||^2 = (\bar{x} - \bar{\mu}_j)^T (\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T \bar{x} - 2\bar{x}^T \bar{\mu}_j + \mu_j^T \mu_j$$

$$= \bar{x}^T \bar{x} - 2(\bar{x}^T \bar{\mu}_j + \frac{1}{2} \mu_j^T \mu_j)$$

$\bar{x}^T \bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2{,}1 \\ 1{,}9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1{,}5 \\ 2{,}0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~when C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1{,}8 & -0{,}4 \\ -0{,}4 & 1{,}2 \end{bmatrix} = \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix}.$$

Given $\bar{x} = [1{,}6 \;\; 1{,}5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} [1{,}5 \;\; 1{,}4] \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix}$$

$$= -\frac{1}{2} [1{,}07 \;\; 0{,}54] \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix} = \underline{-1{,}1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

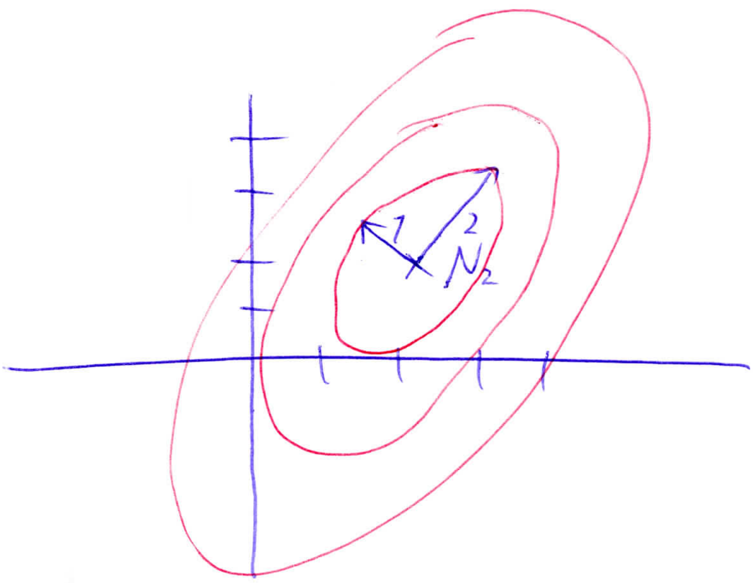
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1$, $\quad \lambda_2 = 2$



Ex 5) We have

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

$$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2{,}2 \end{bmatrix} \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$$g_2(\bar{x}) = -1{,}836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rull to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)\,P(\omega_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}\,|\Sigma_j|^{\frac{1}{2}}}\, e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)\,P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

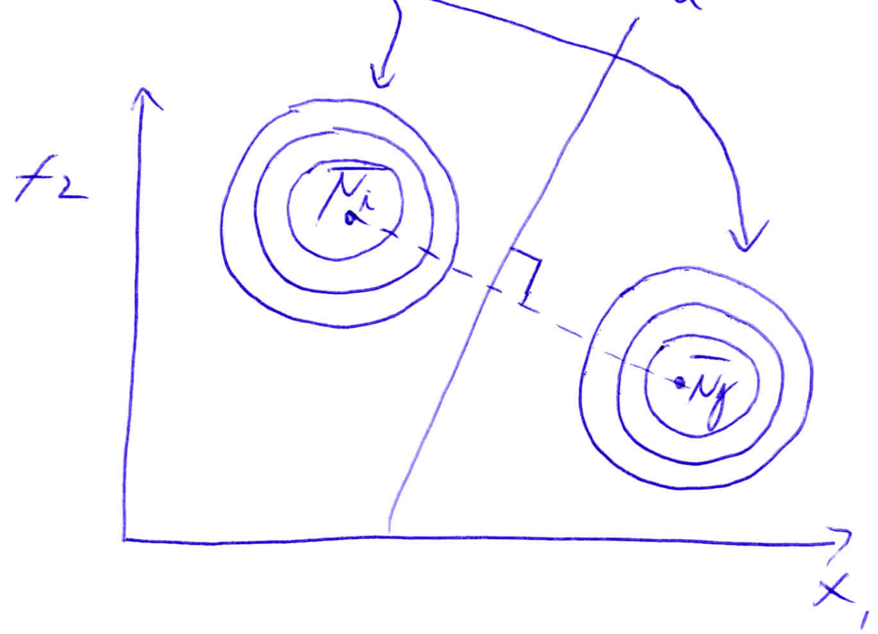
We can discard everything that is common for all classes, so we get:

$$\boxed{g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j)} \quad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.

Eucledian distance



Case 2: Common covariance matrix

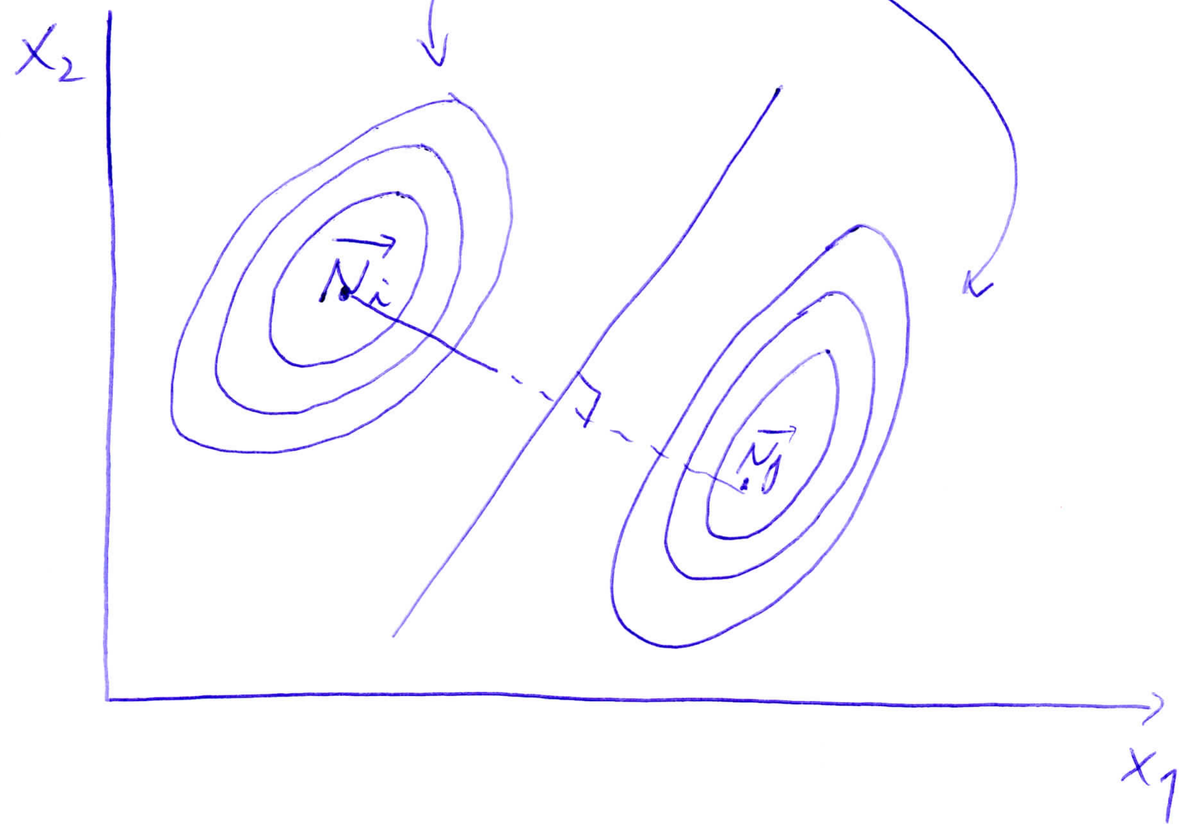
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

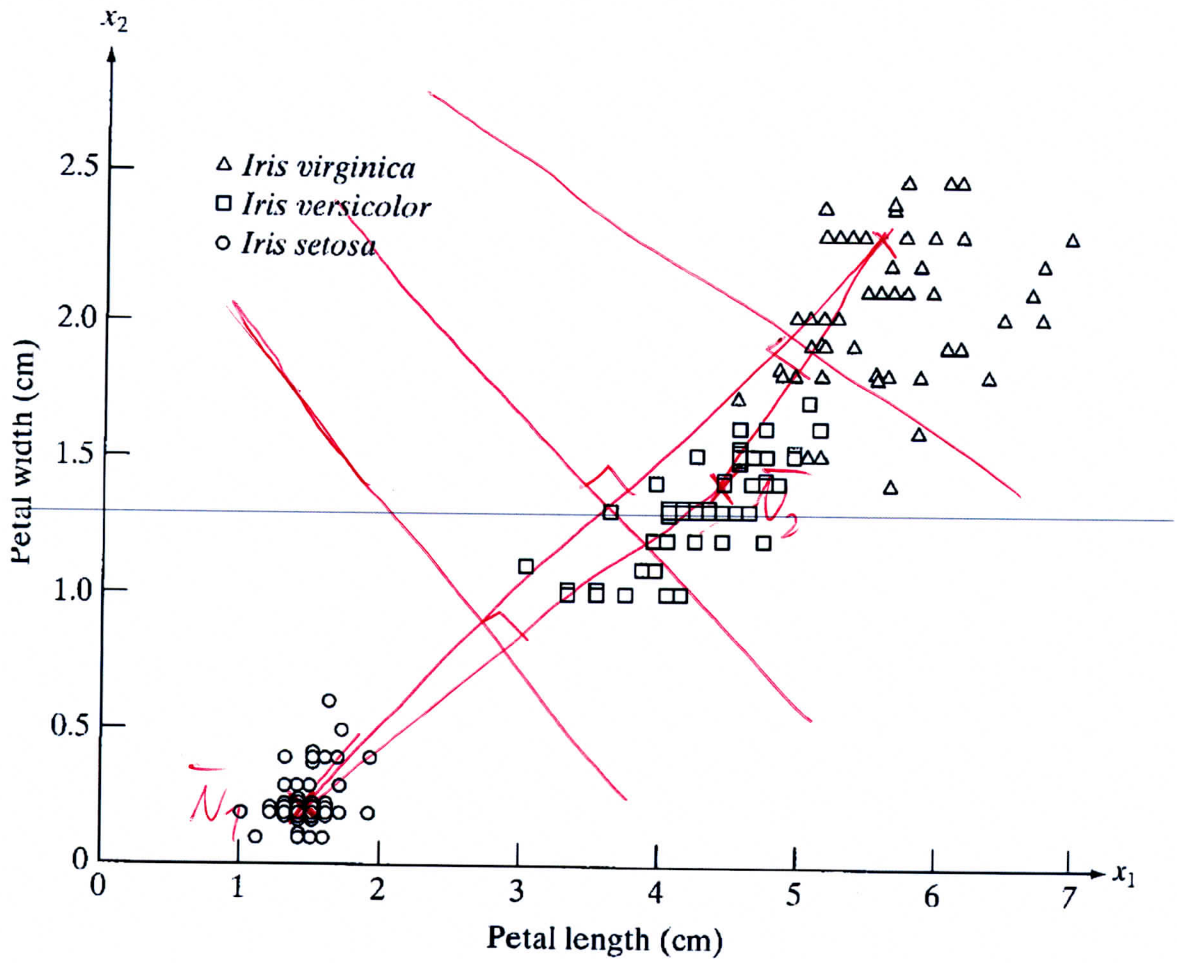
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T (x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T (\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T \bar{x} - 2\bar{x}^T \bar{\mu}_j + \mu_j^T \mu_j$$

$$= \bar{x}^T \bar{x} - 2(\bar{x}^T \bar{\mu}_j + \tfrac{1}{2} \mu_j^T \mu_j)$$

$\bar{x}^T \bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2{,}1 \\ 1{,}9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1{,}5 \\ 2{,}0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1{,}8 & -0{,}4 \\ -0{,}4 & 1{,}2 \end{bmatrix} = \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix}.$$

Given $\bar{x} = [1{,}6 \;\; 1{,}5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} \begin{bmatrix} 1{,}5 & 1{,}4 \end{bmatrix} \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix}$$

$$= -\frac{1}{2} \begin{bmatrix} 1{,}07 & 0{,}54 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix} = \underline{-1{,}1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

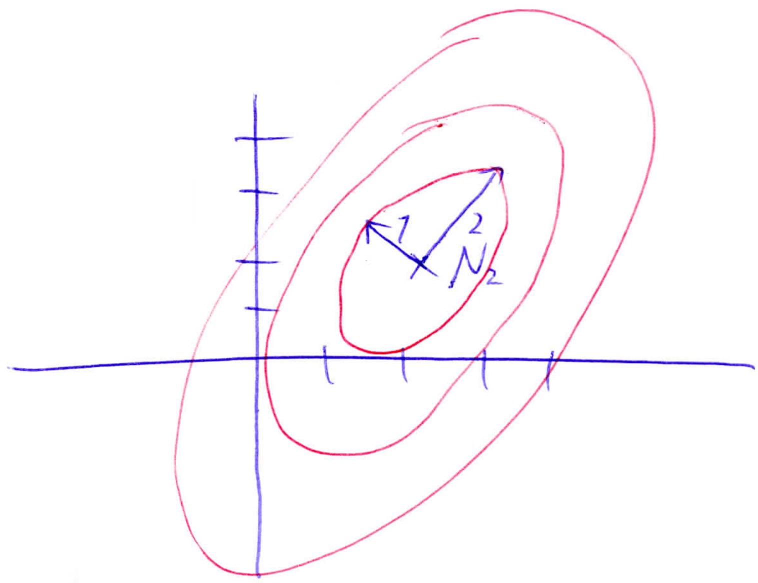
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1 \quad , \quad \lambda_2 = 2$



Ex 5) We have

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

$$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2{,}2 \end{bmatrix} \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$$g_2(\bar{x}) = -1{,}836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rule to find an expression for the class with the highest probability

$$\boxed{p(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|\omega_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|\omega_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}. \text{ And}$$

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalising

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

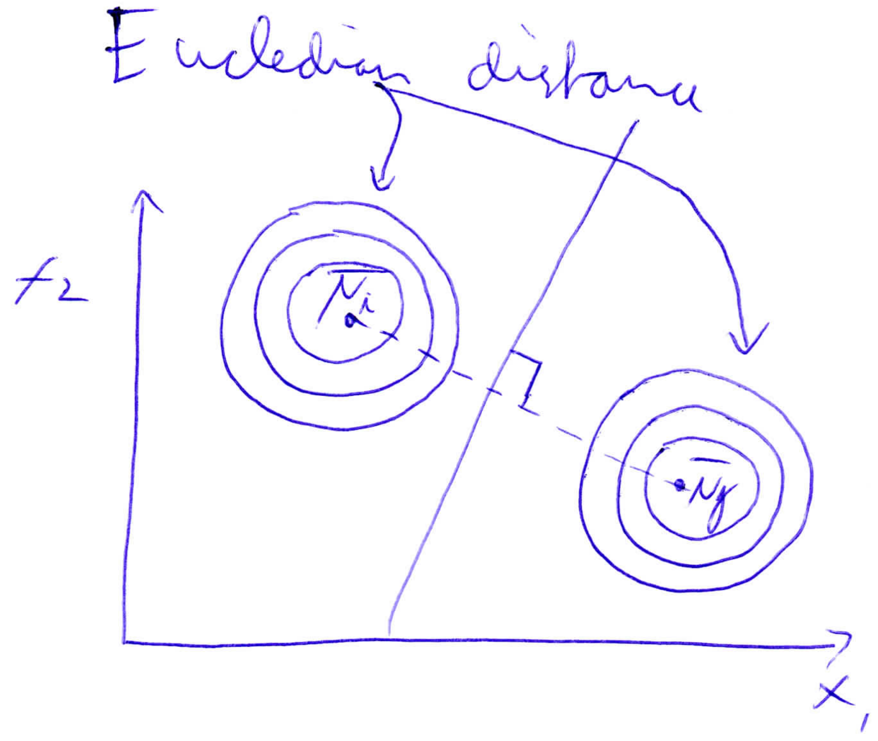
$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

We can discard everything that is common for all classes, so we get:

$$\boxed{g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j)} \tag{2}$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.


Euclidean distance
x_2
μ_i
μ_j
x_1

Case 2: Common covariance matrix

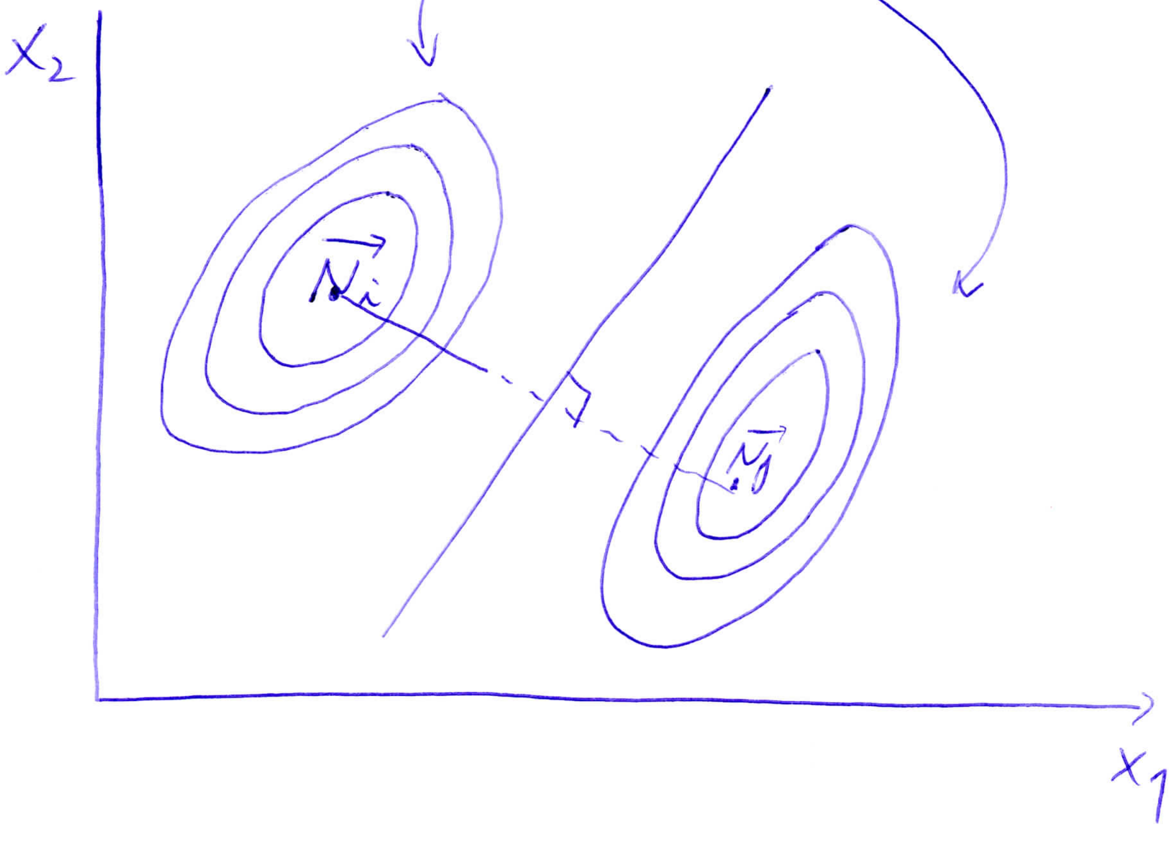
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{N}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{N}_2 = [4{,}3 \quad 1{,}3]^T$

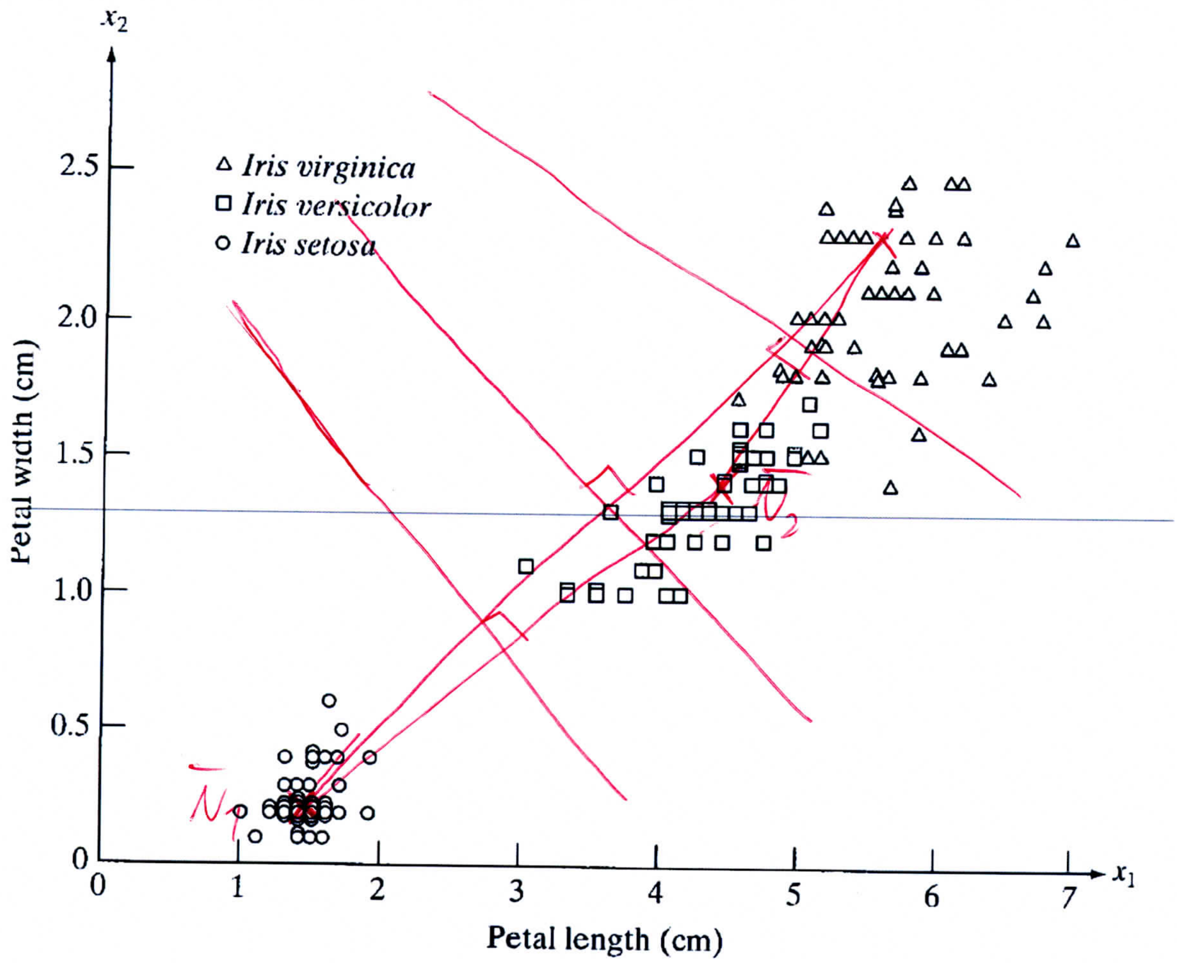
$\bar{N}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T (x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T (\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T \bar{x} - 2\bar{x}^T \bar{\mu}_j + \mu_j^T \mu_j$$

$$= \bar{x}^T \bar{x} - 2\left(\bar{x}^T \bar{\mu}_j + \tfrac{1}{2} \mu_j^T \mu_j\right)$$

$\bar{x}^T \bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T \bar{\mu}_j - \frac{1}{2} \mu_j^T \mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$, ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2}\begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1,5 & 1,4 \end{bmatrix}\begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2}\begin{bmatrix} 1,07 & 0,54 \end{bmatrix}\begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0,1205$

$g_3(\bar{x}) = -4,7095$

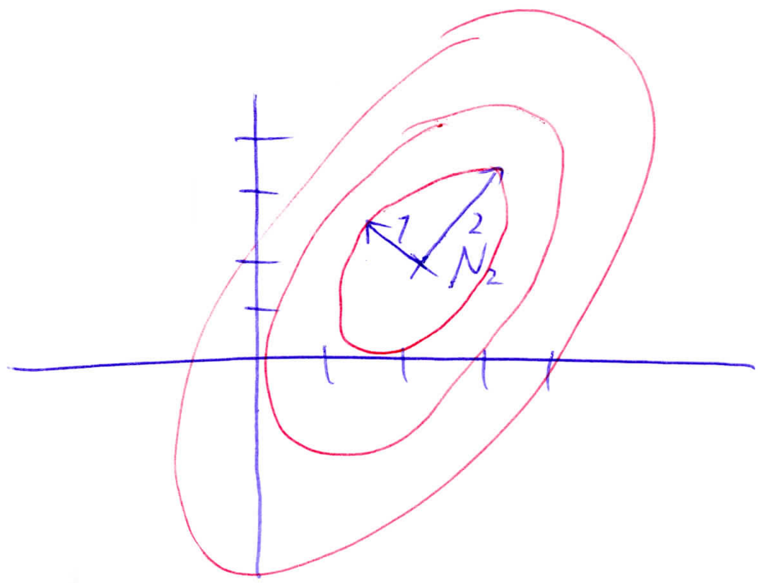
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0,895 \\ 0,447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0,447 \\ 0,895 \end{bmatrix}$

$\lambda_1 = 1 \quad , \quad \lambda_2 = 2$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2} \left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)^T \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \left(\begin{bmatrix} 1,0 \\ 2,2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)$$

$$= -\frac{1}{2} \begin{bmatrix} 1 & 2,2 \end{bmatrix} \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} 0,62 & 1,06 \end{bmatrix} \begin{bmatrix} 1 \\ 2,2 \end{bmatrix}$$

$$= -1,4760$$

$g_2(\bar{x}) = -1,836.$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rull to find an expression for the class with the highest probability

$$\boxed{p(w_j|x) = \frac{p(x|w_j)P(w_j)}{p(x)}} \quad (1)$$

- Any probability can be used to model $p(x|w_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|w_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(w_i|\bar{x}) = \frac{p(\bar{x}|w_j)P(w_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

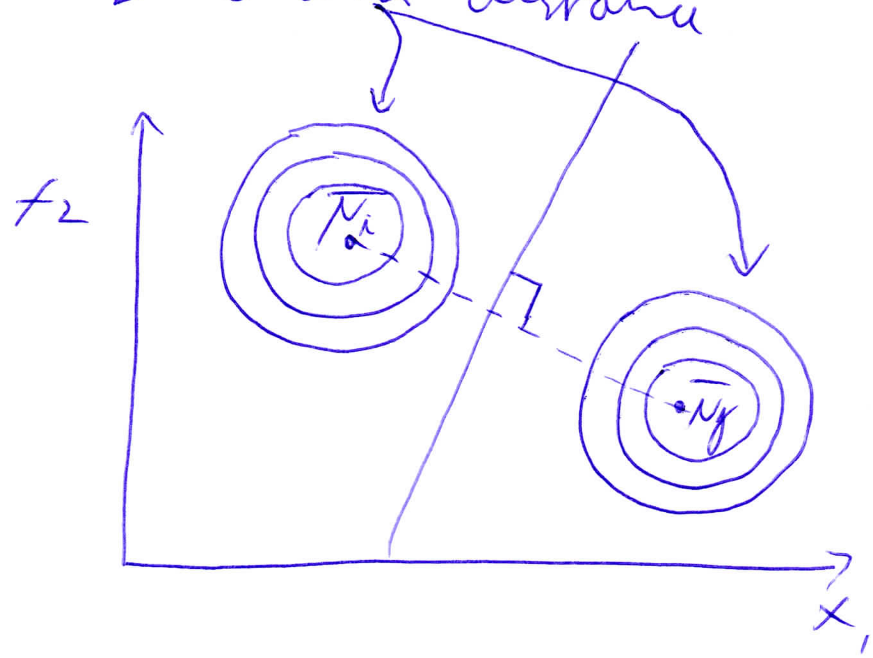
We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \qquad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the <u>Euclidean distance</u>.

This is known as the minimum distance classifier.

Euclidean distance



Case 2: Common covariance matrix

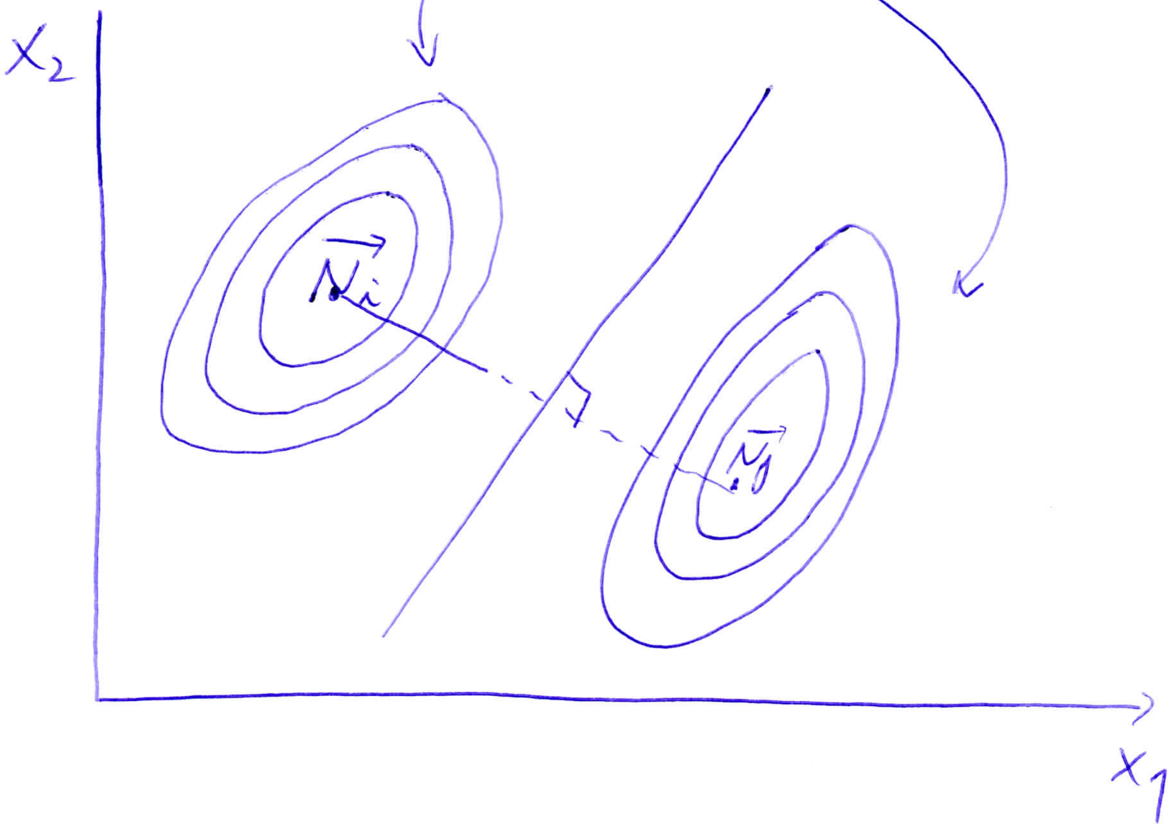
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))}$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\overline{\mu}_1 = [1{,}5 \ \ 0{,}3]^T$

$\overline{\mu}_2 = [4{,}3 \ \ 1{,}3]^T$

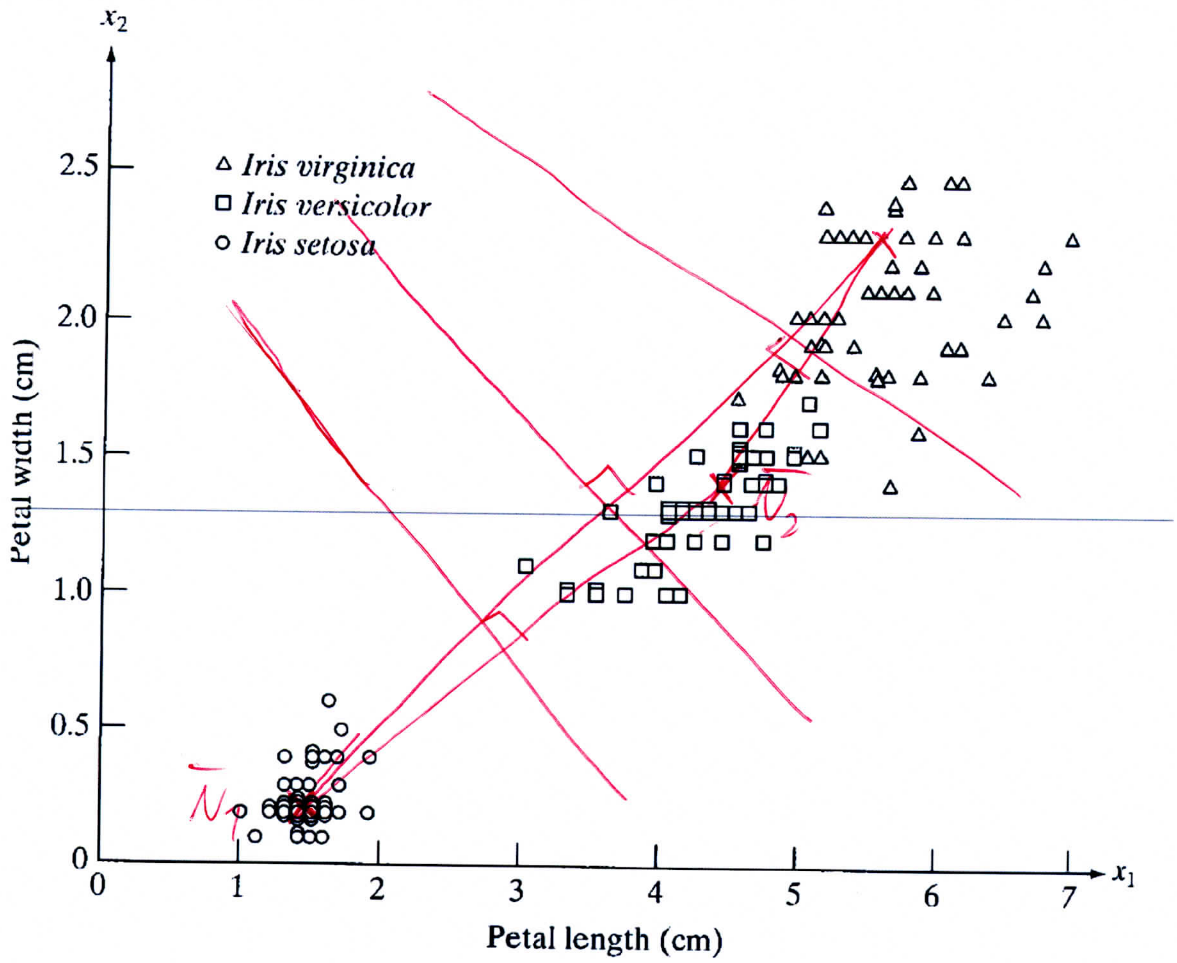
$\overline{\mu}_3 = [5{,}5 \ \ 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$.

So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2\left(\bar{x}^T\bar{\mu}_j + \frac{1}{2}\mu_j^T\mu_j\right)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2,1 \\ 1,9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1,5 \\ 2,0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$. ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1} (\bar{x}-\bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2} \begin{bmatrix} 1,8 & -0,4 \\ -0,4 & 1,2 \end{bmatrix} = \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix}.$$

Given $\bar{x} = [1,6 \;\; 1,5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)^T \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \left(\begin{bmatrix} 1,6 \\ 1,5 \end{bmatrix} - \begin{bmatrix} 0,1 \\ 0,1 \end{bmatrix}\right)$$

$$= -\frac{1}{2} [1,5 \;\; 1,4] \begin{bmatrix} 0,9 & -0,2 \\ -0,2 & 0,6 \end{bmatrix} \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix}$$

$$= -\frac{1}{2} [1,07 \;\; 0,54] \begin{bmatrix} 1,5 \\ 1,4 \end{bmatrix} = \underline{-1,1805}$$

$g_2(\bar{x}) = -0{,}1205$

$g_3(\bar{x}) = -4{,}7095$

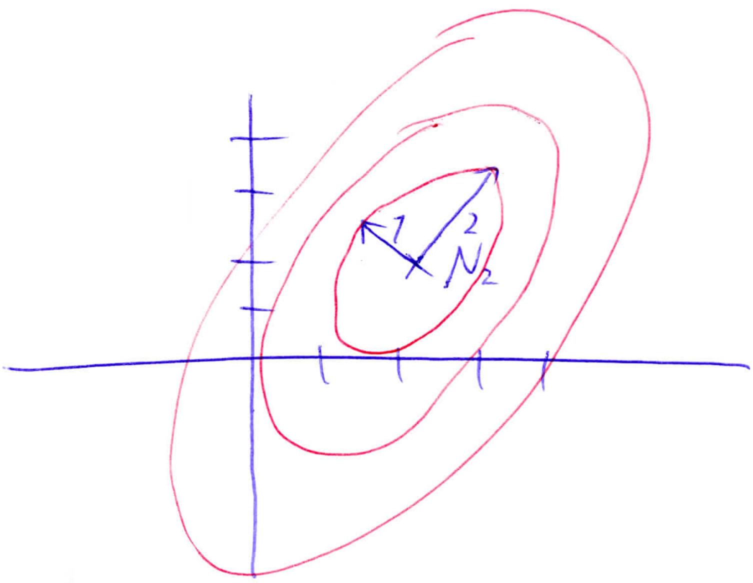
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0{,}895 \\ 0{,}447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0{,}447 \\ 0{,}895 \end{bmatrix}$

$\lambda_1 = 1 \quad , \quad \lambda_2 = 2$



Ex 5) We have

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

$$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2{,}2 \end{bmatrix}\begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix}\begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$$g_2(\bar{x}) = -1{,}836.$$

We choose the maximum, so this is class 1.

MULTIVARIATE GAUSSIAN CLASSIFIER

- We can use Bayes rull to find an expression for the class with the highest probability

$$\boxed{p(w_j|x) = \frac{p(x|w_j)P(w_j)}{p(x)}} \qquad (1)$$

- Any probability can be used to model $p(x|w_j)$, but we use the multivariate Gaussian density:

$$p(\bar{x}|w_j) = \frac{1}{(2\pi)^{n/2}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T\Sigma_j^{-1}(\bar{x}-\bar{\mu}_j)}$$

where $\bar{\mu}_j$ is the mean vector for class j for n features. Giving

$$\bar{\mu}_j = \begin{bmatrix} \mu_j^1 \\ \mu_j^2 \\ \vdots \\ \mu_j^n \end{bmatrix}.$$ And

Σ_j is the covariance matrix for ~~feature~~ class j for n features.

- From (1) we get the discriminant function

$$g_j(x) = P(w_i|\bar{x}) = \frac{p(\bar{x}|w_j)P(w_j)}{p(x)}.$$

But we can ignore $p(x)$ since this is only a normalizing

factor, and use the discriminant function

$$\boxed{g_j(\bar{x}) = p(\bar{x}|\omega_j)P(\omega_j)}.$$

$$g_j(\bar{x}) = \ln(p(\bar{x}|\omega_j)P(\omega_j)) = \ln(p(\bar{x}|\omega_j)) + \ln(P(\omega_j)).$$

$$\boxed{\ln(N \cdot M) = \ln(N) + \ln(M)}$$

If we plug in the multivariate Gaussian classifier, we get:

$$\boxed{g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))}$$

Case 1: $\Sigma_j = \sigma^2 I$

- The features are uncorrelated (independent) and have the same variance.

$$g_j(x) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln P(\omega_j)$$

Let $\Sigma_j = \sigma^2 I$, giving us

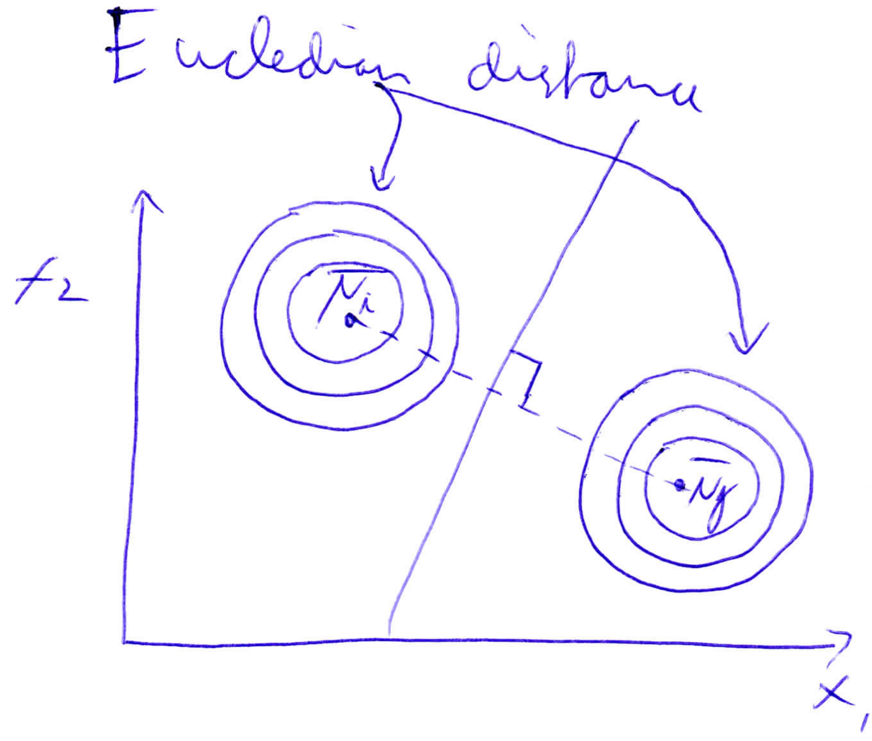
$$g_j(x) = -\frac{1}{2\sigma^2 I}(\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j) - \frac{1}{2}\ln|\sigma^2 I| + \ln P(\omega_j)$$

We can discard everything that is common for all classes, so we get:

$$g_j(x) = -\frac{\|\bar{x}-\bar{\mu}_j\|^2}{2\sigma^2} + \ln P(\omega_j) \qquad (2)$$

where $\|\bar{x}-\bar{\mu}_j\|^2 = (\bar{x}-\bar{\mu}_j)^T(\bar{x}-\bar{\mu}_j)$ is the Euclidean distance.

This is known as the minimum distance classifier.



Euclidean distance
μ_i
μ_j
x_2
x_1

Case 2: Common covariance matrix

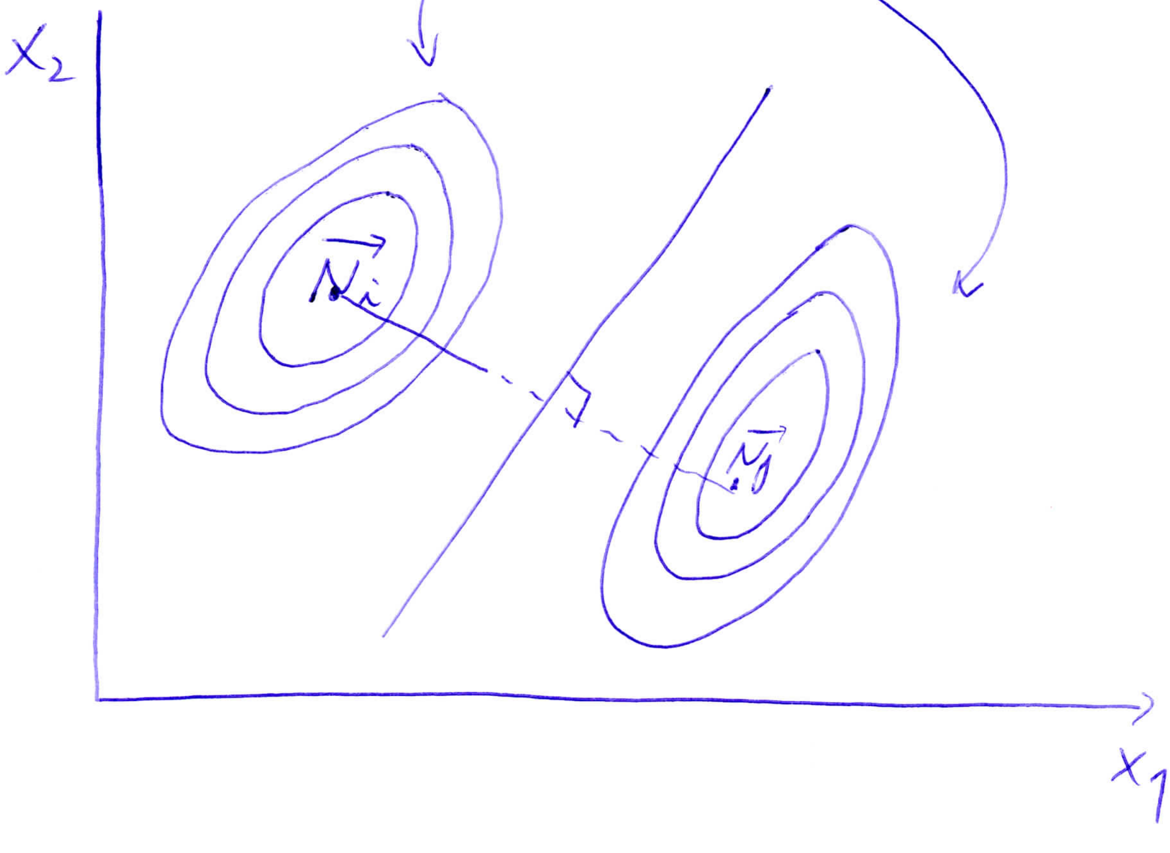
Remember:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma_j^{-1}(\bar{x}-\bar{\mu}_j) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_j| + \ln(P(\omega_j))$$

Since the covariance matrix is equal we get:

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j) + \ln(P(\omega_j))$$

$(\bar{x}-\bar{\mu}_j)^T \Sigma^{-1}(\bar{x}-\bar{\mu}_j)$ is the Mahalanobis distance between $\bar{x}$ and $\bar{\mu}$



$\bar{\mu}_1 = [1{,}5 \quad 0{,}3]^T$

$\bar{\mu}_2 = [4{,}3 \quad 1{,}3]^T$

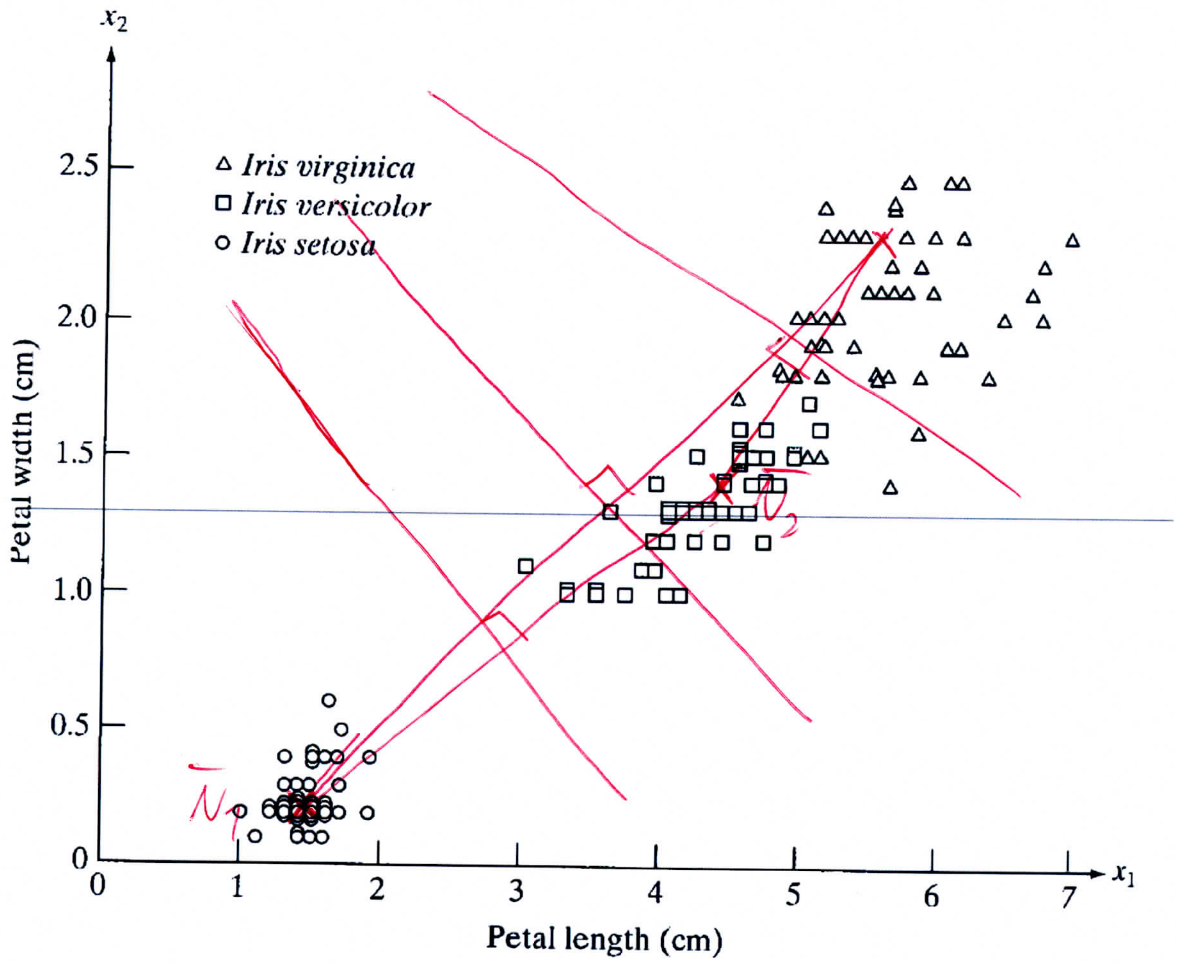
$\bar{\mu}_3 = [5{,}5 \quad 2{,}1]^T$

Solution of selected exercises

Exercises INF 4300 related to the lecture 22.10.14

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.



a. In the above figure, find the class means just by looking at the plot.

b. If this data is classified using a minimum distance classifier, sketch the decision boundaries on the plot.

Solution:

Ex 3) A classifier that uses Euclidean distance computes distance from pattern $\bar{x}$ to class j as:

$$G_j(x) = \|\bar{x} - \bar{\mu}_j\|.$$

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

$$g_j(\bar{x}) = \bar{x}^T \bar{\mu}_j - \frac{1}{2}\mu_j^T \mu_j.$$

Def: $G_j(x) = \|\bar{x} - \bar{\mu}_j\| = \sqrt{(x-\mu_j)^T(x-\mu_j)}$

Since G_j is non-negative, choosing $G_j(x)$ is the same as choosing the smallest $G^2_j(x)$. So we get:

$$G_j^2(x) = \|\bar{x} - \bar{\mu}_j\|^2 = (\bar{x} - \bar{\mu}_j)^T(\bar{x} - \bar{\mu}_j)$$

$$= \bar{x}^T\bar{x} - 2\bar{x}^T\bar{\mu}_j + \mu_j^T\mu_j$$

$$= \bar{x}^T\bar{x} - 2(\bar{x}^T\bar{\mu}_j + \tfrac{1}{2}\mu_j^T\mu_j)$$

$\bar{x}^T\bar{x}$ is independant of j so choosing the minimum of $G^2(x)$ is the same as choosing the maximum of $x^T\bar{\mu}_j - \frac{1}{2}\mu_j^T\mu_j$.

Ex 4) We have

$$\Sigma = \begin{bmatrix} 1{,}2 & 0{,}4 \\ 0{,}4 & 1{,}8 \end{bmatrix}, \quad \bar{\mu}_1 = \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2{,}1 \\ 1{,}9 \end{bmatrix}, \quad \bar{\mu}_3 = \begin{bmatrix} -1{,}5 \\ 2{,}0 \end{bmatrix}$$

We assume that all classes are equally probable, so $P(\omega_j) = \frac{1}{3}$, ~~where C is the~~

a) This is Case 2. Why? $\Sigma_j = \Sigma$

$$g_j(\bar{x}) = -\frac{1}{2}(\bar{x} - \bar{\mu}_j)^T \Sigma^{-1} (\bar{x} - \bar{\mu}_j) + \cancel{\ln(P(\omega_j))}$$

We need Σ^{-1}:

$$\Sigma^{-1} = \frac{1}{2}\begin{bmatrix} 1{,}8 & -0{,}4 \\ -0{,}4 & 1{,}2 \end{bmatrix} = \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix}.$$

Given $\bar{x} = [1{,}6 \;\; 1{,}5]^T$

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \left(\begin{bmatrix} 1{,}6 \\ 1{,}5 \end{bmatrix} - \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1{,}5 & 1{,}4 \end{bmatrix} \begin{bmatrix} 0{,}9 & -0{,}2 \\ -0{,}2 & 0{,}6 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix}$$

$$= -\frac{1}{2}\begin{bmatrix} 1{,}07 & 0{,}54 \end{bmatrix} \begin{bmatrix} 1{,}5 \\ 1{,}4 \end{bmatrix} = \underline{-1{,}1805}$$

$$g_2(\bar{x}) = -0,1205$$

$$g_3(\bar{x}) = -4,7095$$

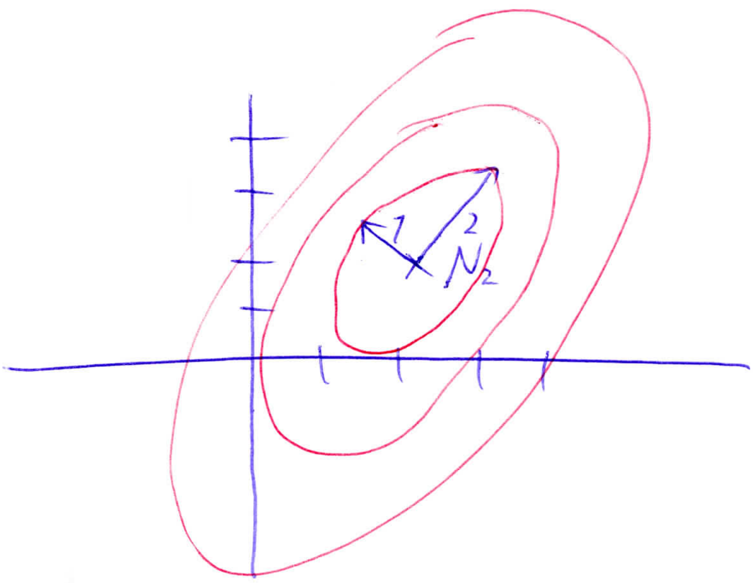
$g_2(\bar{x})$ is the maximum, so we classify $\bar{x}$ as class 2.

b)

$$\Sigma = \begin{bmatrix} 1,2 & 0,4 \\ 0,4 & 1,8 \end{bmatrix}$$

Eigenvectors: $v_1 = \begin{bmatrix} -0,895 \\ 0,447 \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 0,447 \\ 0,895 \end{bmatrix}$

$$\lambda_1 = 1 \quad , \quad \lambda_2 = 2$$



Ex 5) We have

$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$\bar{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\bar{\mu}_2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$\Sigma = \begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix}$, $\Sigma^{-1} = \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}$

Case 2, since $\Sigma_j = \Sigma$.

We are given $x = \begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix}$, this gives us

$$g_1(\bar{x}) = -\frac{1}{2}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)^T \begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\left(\begin{bmatrix} 1{,}0 \\ 2{,}2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$$

$$= -\frac{1}{2}\begin{bmatrix} 1 & 2{,}2 \end{bmatrix}\begin{bmatrix} 0{,}95 & -0{,}15 \\ -0{,}15 & 0{,}55 \end{bmatrix}\begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 0{,}62 & 1{,}06 \end{bmatrix}\begin{bmatrix} 1 \\ 2{,}2 \end{bmatrix}$$

$$= -1{,}4760$$

$g_2(\bar{x}) = -1{,}836.$

We choose the maximum, so this is class 1.