MACHINE LEARNING

EXERCISES

Generative machine-learning models: the Gaussian pattern classifier

All the course material is available on the web site

- Course web site: https://unica-ml.github.io/#

PART 2: Generative machine-learning models: the Gaussian pattern classifier

Required knowledge and problem-solving ability

1. Discriminant functions g(x) for Gaussian classes
   a. General case (different priors, different covariance matrix per class)
   $$g(x;\ \mu, \Sigma) = -\frac{1}{2}x^T\Sigma^{-1}x + \ \mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu + \ln p(\omega) - \frac{1}{2}\ln|\Sigma|$$
   b. Degenerate cases
      i. Same priors, same isotropic covariance matrix $\Sigma = \sigma^2 I$ for each class (Nearest Mean Centroid classifier)
      $g(x) = \mu^T x - \frac{1}{2}\mu^T\mu$ or $g(x) = -||x-\mu||^2$
      ii. Different priors, same isotropic covariance matrix $\Sigma = \sigma^2 I$ for each class
      $g(x) = \frac{1}{\sigma^2}\ \mu^T x - \frac{1}{2\sigma^2}\mu^T\mu + \ln p(\omega)$ or $g(x) = -\frac{||x-\mu||^2}{2\sigma^2} + \ln p(\omega)$
      iii. Same priors, (arbitrary) covariance matrix equal for each class
      $$g(x) = \ \mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu + \ln p(\omega)$$

   Note that, if Gaussians are isotropic but different for each class, g(x) remains quadratic. The degenerate cases listed above hold only <u>when the covariance matrix is the same for each class</u>!

2. Classify a point using the MAP criterion:
   $\text{argmax}_k\ g_k(x)$

3. Plot decision boundaries among classes in feature space
   a. Determine the values of x^* for which it holds that $g_1(x) = g_2(x)$
   b. Find the subset of points x^* for which g_1 and g_2 are the **dominant** classes, i.e., $g_1(x^*) > g_3(x^*)$, $g_1(x^*) > g_4(x^*)$, etc.
   This subset of points will identify the *active* boundary between class 1 and class 2
   c. Repeat for any other pair of classes (<u>see Exercise 2 of Part 2 as an example</u>)

4. Estimate parameter values (mean and covariance matrices) from data (using MLE)
   a. Mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$
   b. Covariance $\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$
   Note that the product inside the summation is an *outer* product between a column and a row vector of *d* elements, being *d* the number of features, and that the sum is computed over the training samples. Accordingly, the covariance matrix is a *d* x *d* matrix.

   Depending on the assumptions made on the covariance matrix (e.g., equal for all classes, isotropic, etc.), one should adjust the above estimate accordingly (see Exercise 4 for some examples).

Exercise 1

Consider a two-class problem in $\mathbb{R}^2$ (two-dimensional feature space)

Each class has a Gaussian probability density function:

$$p(\boldsymbol{x}|\omega_i) = \frac{1}{(2\,\pi)^{d/2}\,|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu_i})^T \Sigma_i^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu_i})\right]$$

Each pattern is characterized by a numerical feature vector in R^2 : $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Let us denote with $\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i,1} \\ \mu_{i,2} \end{pmatrix}$ the mean vector of the *i*-th class

Using the MAP decision rule:

a) Classify the pattern x_T=(1/2, 1/3)' using the MAP rule, compute explicitly the decision regions and the decision region for the pattern x_T=(1/2, 1/3)').

b) Classify the pattern x_T=(1/2, 1/3)' by the likelihood ratio test.

Use the following data:

$$\mathrm{P}(\omega_1) = \mathrm{P}(\omega_2); \quad \Sigma_1 = \Sigma_1 = \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix};$$

Note: the aim of this exercise is to show that it is (usually) easier to solve a classification problem for a single pattern in R^d computing the likelihood ratio test rather than computing explicitly the decision regions. However, the explicit computation of the decision regions provides additional ("geometrical") information on the decision mechanism of the pattern classifier.

Solution

The distributions of the two classes are:

$$p(\boldsymbol{x}|\omega_i) = \frac{1}{2\,\pi} \exp\left[-\frac{1}{2}\,\|\boldsymbol{x}-\boldsymbol{\mu}_i\|^2\right]$$

$$p(\boldsymbol{x}|\omega_1) = \frac{1}{2\,\pi} \exp\left[-\frac{1}{2}\,\|\boldsymbol{x}-\boldsymbol{\mu}_1\|^2\right] = \frac{1}{2\,\pi} \exp\left[-\frac{1}{2}\,[(x_1)^2 + (x_2)^2]\right]$$

$$p(\boldsymbol{x}|\omega_2) = \frac{1}{2\,\pi} \exp\left[-\frac{1}{2}\,\|\boldsymbol{x}-\boldsymbol{\mu}_2\|^2\right] = \frac{1}{2\,\pi} \exp\left[-\frac{1}{2}\,\left(\left(x_1-\mu_{2,1}\right)^2 + \left(x_2-\mu_{2,2}\right)^2\right)\right] =$$

$$= \frac{1}{2\,\pi} \exp\left[-\frac{1}{2}\,[(x_1)^2 + (x_2)^2 + 2 - 2\,x_1 - 2\,x_2]\right]$$

a) Classify the pattern x_T=(1/2, 1/3)' using the decision regions

Priors (class prior probabilities) can be disregarded, as they are equal.

The decision region R1 for **class 1** is defined by

$p(\boldsymbol{x}|\omega_1) > p(\boldsymbol{x}|\omega_2)$

that is,

$$\exp\left[-\frac{1}{2}\left[(x_1)^2+(x_2)^2\right]\right] > \exp\left[-\frac{1}{2}\left[(x_1)^2+(x_2)^2\right]\right]\ \exp\left[-\frac{1}{2}\left[2-2\,x_1-2\,x_2\right]\right]$$

$$1 > \exp\left[-\frac{1}{2}\left[2-2\,x_1-2\,x_2\right]\right]$$

Applying the logarithm:

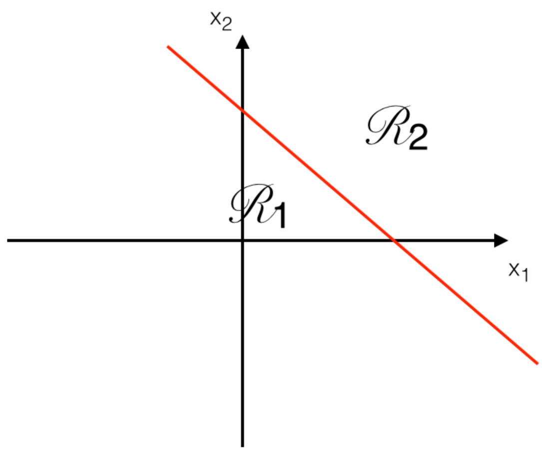
$$0 > x_1 + x_2 - 1 \Leftrightarrow$$

$$x_1 + x_2 < 1 \Leftrightarrow$$

$$x_2 < 1 - x_1$$

Therefore, the two decision regions R1 and R2 are:

$$\Re_1 = \{\mathbf{x} \mid x_1 + x_2 < 1\}$$
$$\Re_2 = \{\mathbf{x} \mid x_1 + x_2 > 1\}$$



The patterns for which $x_1 + x_2 = 1$ (decision boundary) can be assigned randomly to class 1 or class 2.

It is easy to see that pattern x_T=(1/2, 1/3)' belongs to the first region, $x_T \in \Re_1$, and thus it is assigned to class 1.

It is easy to see that in this particular exercise would be possible to find the correct solution using symmetry considerations only. Please, try to draw the two class-related distributions and check that that it was possible to spot the right solution.

It is worth noting that the decision boundary is very simple in this case. Both distributions are Gaussian and the covariance matrix is an identity matrix. Indeed, we know that the decision boundary is linear for such distributions and this is the Nearest Mean Centroid classifier.

In other cases, it is almost impossible to obtain an explicit analytical condition from the inequality:

$$P(\omega_1)p(\mathbf{x}/\omega_1) > p(\mathbf{x}/\omega_2)P(\omega_2)$$

And the related decision regions are much more complex (non linear decision boundaries).

b)classify the pattern x_T=(1/2, 1/3)' using the likelihood ratio test

The pattern is assigned to class 1 if:

$$l(x) = \frac{p(\mathbf{x}/\omega_1)}{p(\mathbf{x}/\omega_2)} > \frac{P_2}{P_1} = \theta,$$

with $\theta = 1$ given that the priors are equal.

$$l(\boldsymbol{x}) = \frac{\frac{1}{2\pi}\exp\left[-\frac{1}{2}\left((x_1)^2+(x_2)^2\right)\right]}{\frac{1}{2\pi}\exp\left[-\frac{1}{2}\left((x_1)^2+(x_2)^2-2x_1-2x_2+2\right)\right]} = \frac{1}{\exp\left[-\frac{1}{2}(2-2x_1-2x_2)\right]} = \exp(1-x_1-x_2)$$

$$l(\mathrm{x}) = \exp[\,1 - x_1 - x_2]$$

Class 1: $l(\mathrm{x}) > 1$

Class 2: $l(\mathrm{x}) < 1$

Therefore:

$$l(\mathbf{x}_T) = \exp[1-(1/2+1/3)] = \exp(1/6) \cong 1.184 > 1$$

The pattern x_T belongs to the class 1.

It is worth to note that the likelihood ratio test always involves scalar values. Furthermore, it is not necessary to simplify the analytical expression of *l*(x), as we did above. It is sufficient to compute the actual value of *l*(x) and compare its value with the threshold.

Note: here we applied the MAP rule step by step, to explain how to compute the decision regions and how to apply the likelihood ratio test. However, if the considered distributions are the Gaussian (as in this case), we know the forms of the discriminant functions and can compute them directly (see Part 2 of the course for classifiers based on the Gaussian model).

Exercise 2

Let us consider a **three-class** problem, with a **bi-dimensional feature space**.

a. Classify the pattern $x_t = (1,1/3)^T$ under the following assumptions:

- the probability density function of each class is Gaussian, with the same covariance matrix $\mathbf{\Sigma}=\sigma^2\mathbf{I}$, the parameter σ^2 (variance) is **unknown**
- all the classes have the same prior probability
- the mean vectors of the three classes are: $\mu_1=(0,0)^T$; $\mu_2=(0,2)^T$; $\mu_3=(2,1)^T$.

b. Classify the pattern $x_t =(1,1/3)^T$ with these assumptions about priors (class prior probabilities):

$P(\omega_1) = P(\omega_2) = 1/4;\ P(\omega_3) = 1/2$

Note: the goal of this exercise is to point out the roles played by class variance and class prior probabilities. When all the classes have the same prior probability, variance does not matter for pattern classification. If the priors are different, variance matters and can counterbalance differences of class prior probabilities. See the equation of the decision boundary (x_0) for one-dimension feature space (Parte 2 of the course):

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

Solution

We know from Part 2 that the general form of a multivariate normal (Gaussian) pdf (pdf states for Probability Density Function) is:

$$p(x|\omega_i) = \frac{1}{(2\,\pi)^{\frac{d}{2}}\,|\Sigma_i|^{\frac{1}{2}}} \exp[-\frac{1}{2}\,(x-\mu_i)^T\,\Sigma_i^{-1}\,(x-\mu_i)]$$

The discriminant function $g_i(x)$ is (see Part 2 of the course):

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\,\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\left|\mathbf{\Sigma}_i\right| + \ln P(\omega_i)$$

Considering a Gaussian model with $\Sigma_i=\sigma^2 I$ we have (see Part 2 of the course):

$$\left|\mathbf{\Sigma}_i\right| = \sigma^{2d}\,;\ \mathbf{\Sigma}_i^{-1} = \left(1/\sigma^2\right)\mathbf{I}$$

and we obtain the discriminant functions:

$$g_i(\mathbf{x}) = -\frac{\left\|\mathbf{x}-\boldsymbol{\mu}_i\right\|^2}{2\sigma^2} + \ln\left(P(\omega_i)\right)$$

Note: the above equation makes clear the roles played by class variance and class prior probabilities. When all the classes have the same prior probability and the same variance, variance does not matter for pattern classification. If the priors are different, variance matters and can counterbalance class prior probabilities; the value of priors is all more important as the variance is large, and vice versa.

Try to draw simple examples of Gaussian distributions with different priors and different variance values in order to understand the roles played by class variance and class prior probabilities.

a) Equal priors

For this particular case, the discriminant functions become:

$g_i(\mathbf{x}) = -\|\mathbf{x} - \boldsymbol{\mu}_i\|^2$ (variance and priors can be disregarded as they have same values for all the classes)

for the pattern $x_t=(1, 1/3)^T$, the values of the three discriminant functions are:

$$g_1(x)|_{x=x_t} = -\|x - \mu_1\|^2 = -\left[(1-0)^2 + \left(\frac{1}{3} - 0\right)^2\right] = -\left(1 + \frac{1}{9}\right) = -10/9$$

$$g_2(x)|_{x=x_t} = -\|x - \mu_2\|^2 = -\left[(1-0)^2 + \left(\frac{1}{3} - 2\right)^2\right] = -\left(1 + \frac{25}{9}\right) = -34/9$$

$$g_3(x)|_{x=x_t} = -\|x - \mu_3\|^2 = -13/9$$

Comparing the values of the above discriminant functions g_1, g_2 and g_3 for the pattern x_t, we see that the value of the posterior probability is maximum for class 1. Therefore, the pattern $(1,1/3)^T$ is assigned to class 1, **regardless of the unknown value of $\boldsymbol{\sigma}$.** This is a result that we could guess easily (see Part 2 of the course), when we have same priors, same isotropic covariance matrix $\Sigma = \sigma^2 \mathrm{I}$ for each class, we have the **Nearest Mean Centroid classifier**. The pattern $(1,1/3)^T$ is closer to the centroid of class 1.

b) **Case with different priors**

$$P(\omega_1) = P(\omega_2) = 1/4;\ P(\omega_3) = 1/2$$

Note: prior probability of class 3 is double of the priors of the other classes. On the other hand, the pattern $(1,1/3)^T$ is closer to the centroid of class 1. We can expect that a larger prior is dominant when the variance becomes large, namely, a large variance can counterbalance the "distance" term (distance from the centroid of the class) and makes class prior probability more important.

For this particular case, the linear discriminant function is:

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x} - \boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln(P(\omega_i))$$

for the pattern $x_t=(1, 1/3)^T$, we have:

$$g_1(\mathbf{x}) = \frac{-\|\mathbf{x} - \boldsymbol{\mu}_1\|^2}{2\sigma^2} + \ln(1/4)$$

$$g_2(\mathbf{x}) = \frac{-\|\mathbf{x}-\boldsymbol{\mu}_2\|^2}{2\sigma^2} + \ln(1/4)$$

$$g_3(\mathbf{x}) = \frac{-\|\mathbf{x}-\boldsymbol{\mu}_3\|^2}{2\sigma^2} + \ln(1/2)$$

Note that the inequality $g_1(x)>g_2(x)$ does not depend on the value of ln(1/ 4), whereas the other two inequalities depend on the difference between ln(1/4) and ln(1/2).

Let's calculate the value of the discriminant functions **for the considered pattern x=x$_t$:**

$$g_1(\mathbf{x}) = \frac{-\|\mathbf{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln(1/4) = -\frac{10/9}{2\sigma^2} + \ln(1/4)$$

$$g_2(\mathbf{x}) = \frac{-\|\mathbf{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln(1/4) = -\frac{34/9}{2\sigma^2} + \ln(1/4)$$

$$g_3(\mathbf{x}) = \frac{-\|\mathbf{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln(1/2) = -\frac{13/9}{2\sigma^2} + \ln(1/2)$$

Now let's compare the discriminant functions for the three classes (Part 2 of the course), in order to find the function for which we have the maximum value for the pattern $(1,1/3)^T$. We compare each class with respect to all the other classes in order to find the "maximum".

Class 1 vs class 2

For the considered pattern, the inequality $g_1(x_t)>g_2(x_t)$ is always satisfied.

Class 1 vs class 3

From the inequality $g_1(x_t)>g_3(x_t)$, we obtain:

$$-\frac{10/9}{2\sigma^2} + \ln(1/4) > -\frac{13/9}{2\sigma^2} + \ln(1/2)$$

$$\sigma^2 < \frac{3}{18}\left(\frac{1}{\ln(4)-\ln(2)}\right) = 0.2404$$

Pattern $(1,1/3)^T$ is assigned to class 1 if $\sigma^2 < 0.2404$

Pattern $(1,1/3)^T$ is assigned to class 3 if $\sigma^2 > 0.2404$

Class 2 vs. class 3

From the inequality $g_2(x_t)>g_3(x_t)$ we obtain

$$\log(2) - \log(4) > \frac{7}{6}\frac{1}{\sigma^2}$$

This inequality is never satisfied, accordingly, the classification rule is:

Pattern $(1,1/3)^T$ is assigned to class 1 if $\sigma^2 < 0.2404$

Pattern $(1,1/3)^T$ will be assigned to class 3 if $\sigma^2 > 0.2404$

Note that the result is just the expected result! The larger prior for class 3 is dominant when the variance is larger than the value 0.2404, namely, a large variance can counterbalance the "distance" term (distance from the centroid of the class) and makes class prior probability more important. You should remember that the pattern $(1,1/3)^T$ is closer to the centroid of class 1, but this "distance" term is less important when the variance is larger.

Homework:

Find and draw the decision boundaries for the following three cases:

a) equal priors
b) P1=P2=1/4, $\sigma = 0.1$
c) P1=P2=1/4, $\sigma = 0.5$

Exercise 3

Let us consider a hypothetical problem of detection of network intrusions (a computer security problem). We want to classify network traffic using two characteristic measures (two features). We have three data classes of interest (denial-of-service attack, probing attack, and normal traffic). Traffic patterns follow a normal distribution with different, known parameters (i.e., parameters have been previously estimated using the Maximum Likelihood method, see Part 2 of the course).

For the sake of simplicity, let us assume that covariance matrices are equal for the three classes:

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix},$$

And the centroids (mean values) of the three classes are:

$$\mathbf{m}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \mathbf{m}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \ \mathbf{m}_3 = \begin{bmatrix} -2 \\ -2 \end{bmatrix},$$

Moreover, let us suppose that class prior probabilities are as follows (they have been estimated by a training set):

$$P(\omega_1) = P(\omega_2) = \frac{1}{4}$$

A)Find the discriminant functions using the MAP decision rule**.** Classify the network traffic pattern characterized with the vector $x_c=(1,1)^t$.

B)Find the decision boundaries.

SOLUTION

The general form of a multivariate normal p.d.f. (probability density function) is:

$$p(x|\omega_i) = \frac{1}{(2\,\pi)^{\frac{d}{2}}\,|\Sigma_i|^{\frac{1}{2}}} \exp[-\frac{1}{2}\,(x-\mu_i)^T\,\Sigma_i^{-1}\,(x-\mu_i)]$$

The discriminant function $g_i(x)$ is:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\,\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

A)Find the discriminant functions with the MAP decision rule**.** Classify the network traffic pattern $x_c=(1,1)^t$.

From Part 2 of the course, we know that if $\Sigma_i=\Sigma$ then we have (in our feature space) hyper-ellipsoidal "clusters" with same size, shape, centroids μ_i . According to the MAP rule, we know that the discriminant functions can be written as follows:

$$g_i(x) = \left[-\frac{1}{2} \cdot (x - m_i)^T \cdot \Sigma^{-1} \cdot (x - m_i) \right] + \ln(p(\omega_i))$$

$$= -\frac{1}{2}\left[\underbrace{\boldsymbol{x}^T\,\Sigma^{-1}\,\boldsymbol{x}} + \underbrace{(-\boldsymbol{x}^T\,\Sigma^{-1}\,\boldsymbol{\mu_i} - \boldsymbol{\mu_i}^T\,\Sigma^{-1}\,\boldsymbol{x})} + \boldsymbol{\mu_i}^{\,T}\,\Sigma^{-1}\,\boldsymbol{\mu_i}\right] + \ln p(\omega_i) \Rightarrow$$

$$\Rightarrow g_i(x) = -\frac{1}{2}\left[-2\,\boldsymbol{\mu_i}^T\,\Sigma^{-1}\,\boldsymbol{x} + \boldsymbol{\mu_i}^{\,T}\,\Sigma^{-1}\,\boldsymbol{\mu_i}\right] + \ln p(\omega_i) =$$

$$= \boldsymbol{\mu_i}^T\,\Sigma^{-1}\,\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu_i}^{\,T}\,\Sigma^{-1}\,\boldsymbol{\mu_i} + \ln p(\omega_i)$$

We know that:

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}, |\Sigma| = 7 \qquad \Sigma^{-1} = \tfrac{1}{7}\begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$$

Therefore, for each class, we can write the discriminant functions as follows (remember that P_1=P_2=1/4 and, consequently, P_2=1/2):

$g_1(\boldsymbol{x}) = \ln p(\omega_1) = \ln\frac{1}{4}$ (remember that centroid of class 1 is “zero”)

$$g_2(\boldsymbol{x}) = (2\ \ 2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 4\ + \ln p(\omega_2)\ = 2(x_1 + x_2) - 4\ + \ln\frac{1}{4}$$

Basic calculations for g2(x):

$$\bar{\mu}_1^T\,\Sigma^{-1}\,\bar{x} \Rightarrow (2,2)\,\frac{1}{7}\begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$$

$$\frac{1}{7}\,(14,14)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (2,2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$g_3(\boldsymbol{x}) = -(2\ \ 2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 4\ + \ln p(\omega_3)\ = -2(x_1 + x_2) - 4\ + \ln\frac{1}{2}$$

As expected, as all the covariance matrices are equal, we obtain **linear discriminant functions**. Decision boundaries are defined by planes placed between each “neighboring” class.

Let us classify pattern $x_c=(1,1)^t$.

$g_1(x_c)$=-1.3863
$g_2(x_c)$=-1.3863
$g_3(x_c)$=-8.6931

Thus, such a point can be assigned to class 1 or 2.

B) Find the decision boundaries

Graphs shown below are not strictly necessary for the solution of the problem; they are shown for the sake of clarity.

The eigenvectors of Σ are as follows: $\begin{pmatrix} -\sqrt{2}/2 \\ -\sqrt{2}/2 \end{pmatrix}$ and $\begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}$, with eigenvalues 1 and 7, respectively;

The Gaussian distributions of the three classes are depicted below:

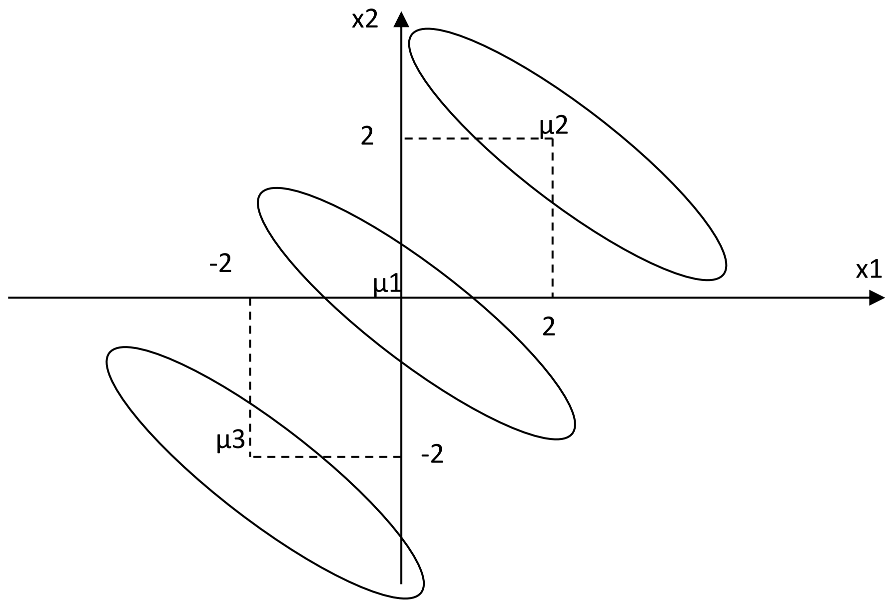


Fig.1 Probability density functions of the three classes (schematic representation of contour lines)- prior probabilities are not considered

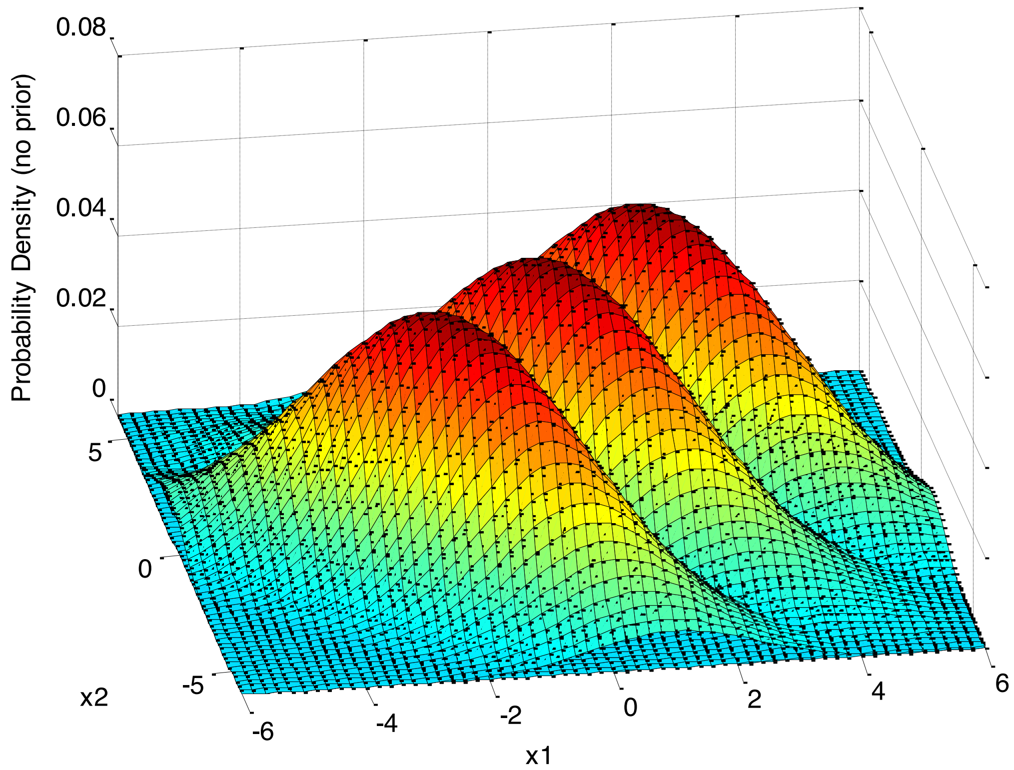


Fig.2 Probability density functions of the three classes - prior probabilities are not considered

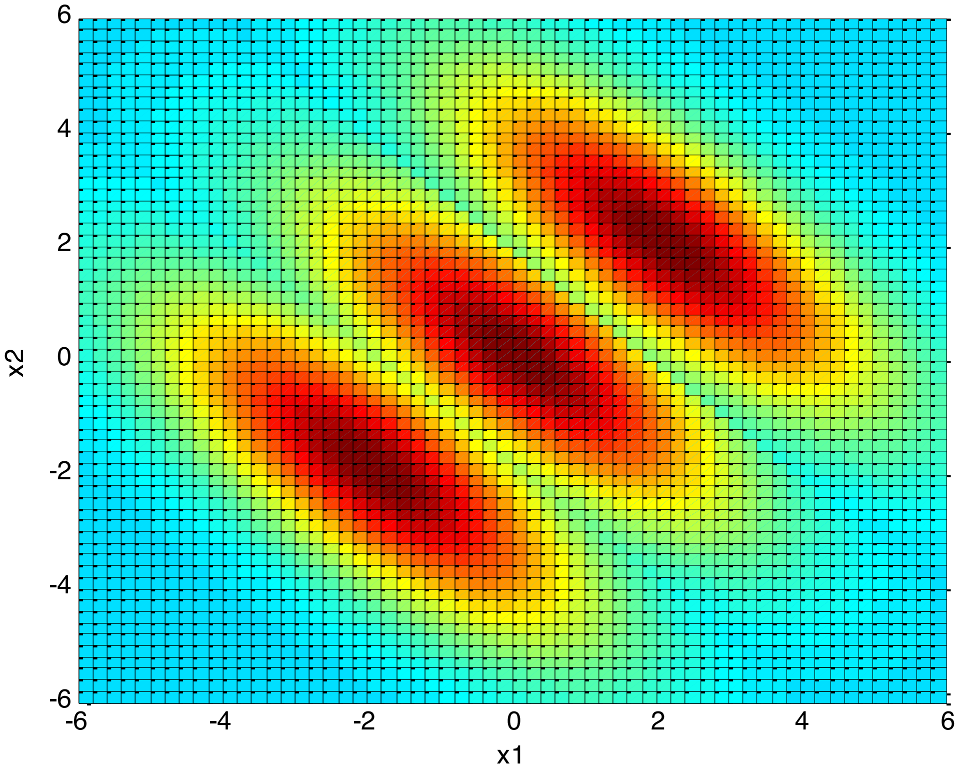


Fig.3 Probability density functions of the three classes - prior probabilities are not considered

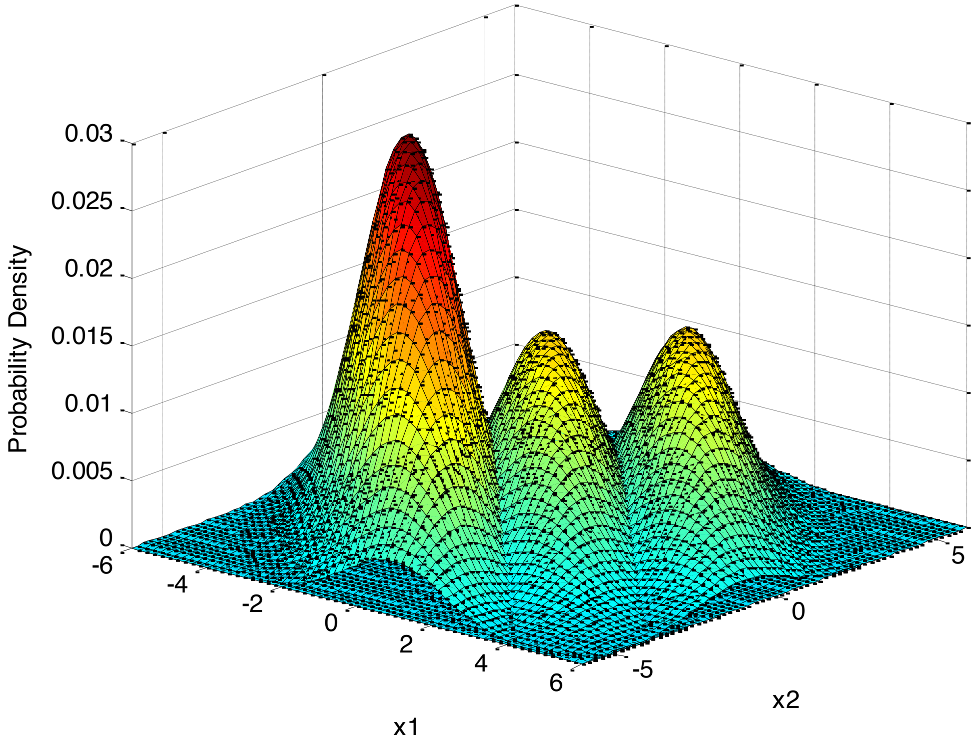


Fig.4 Probability density functions of the three classes - prior probabilities are considered. 1/4 (class 1), 1/4 (class 2), 1/2 (class 3)

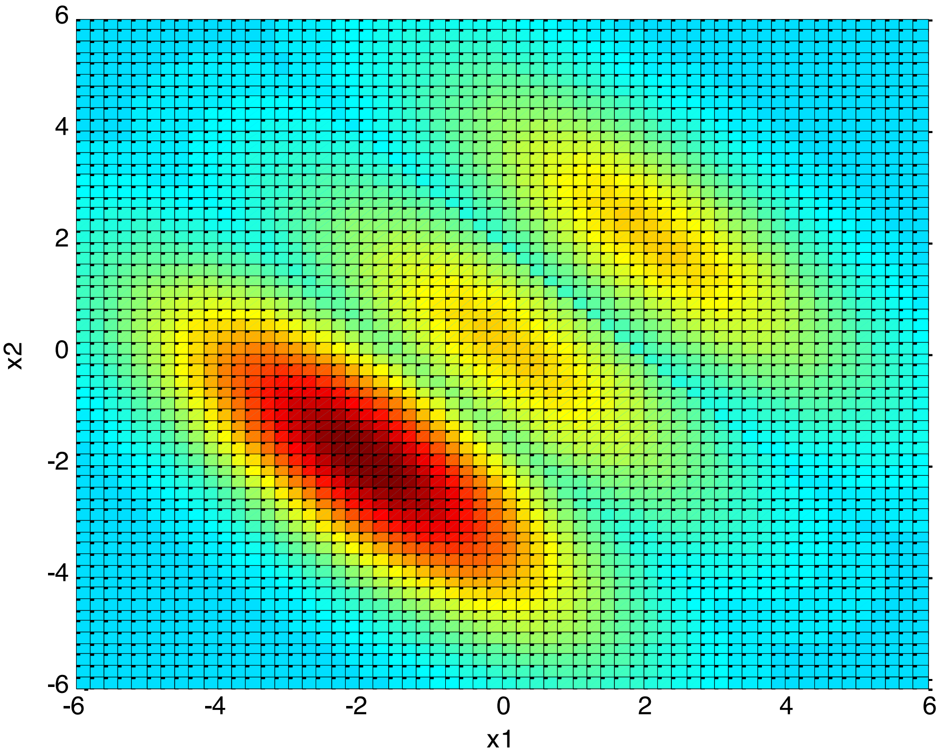


Fig.4 Probability density functions of the three classes - prior probabilities are considered. 1/4 (class 1), 1/4 (class 2), 1/2 (class 3)

We know from Part 2 that decision boundaries correspond to: $g_i(x) = g_j(x)$, that is, to the set of points for which the posterior probability for class *i* and class *j* are identical, provided that no other class has higher probability.

We already computed the discriminant functions, they are:

$$g_1(\boldsymbol{x}) = \ln\frac{1}{4}$$

$$g_2(\boldsymbol{x}) = 2(x_1 + x_2) - 4 \ + \ln\frac{1}{4}$$

$$g_3(\boldsymbol{x}) = -2(x_1 + x_2) - 4 \ + \ln\frac{1}{2}$$

First decision boundary (between class 1 and class 2)**:**

$$g_1(\boldsymbol{x}) = g_2(\boldsymbol{x}) \Longrightarrow 2(x_1 + x_2) - 4 = 0$$

$$(x_1 + x_2) = 2$$

$\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}$ is defined by equation $x_1 + x_2 = 2$

Now, we should check if this is a proper decision boundary between class 1 and 2.
We can check that by substituting the equation of the above boundary into the discriminant functions g1(), g2 ( ) and g3 ( ) and check the value of these functions for this boundary defined by g1( )= g2 ( ).
We find that:
$g_1(\hat{\mathbf{x}}) = -1.3863;$

$g_2(\hat{\mathbf{x}}) = -1.3863;$

$g_3(\hat{\mathbf{x}}) = -8.6931;$

Obviously, we find the equality $g_1(\hat{\mathbf{x}}) = g_2(\hat{\mathbf{x}})$.
The inequality $g_3(\hat{\mathbf{x}}) < g_1(\hat{\mathbf{x}}) = g_2(\hat{\mathbf{x}})$ tells us that this is a decision boundary between class 1 and 2, for all values of $\hat{x}$.

Second decision boundary (between class 2 and class 3)**:**

$$g_2(\boldsymbol{x}) = g_3(\boldsymbol{x}) \Longrightarrow 2(x_1 + x_2) \; + \ln\frac{1}{4} = -2(x_1 + x_2) + \ln\frac{1}{2}$$

$$\Rightarrow (x_1 + x_2) = \frac{1}{4}(\ln 2)$$

$\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}$ is defined by equation $x_1 + x_2 = \frac{1}{4}\ln 2$

Substituting this equation into the discriminant functions g1( ), g2 ( ) and g3 ( ), we obtain the value of these functions when g2( )= g3 ( ) :

$$g_1(\hat{\mathbf{x}}) = -1.3863;$$
$$g_2(\hat{\mathbf{x}}) = -5.0397;$$
$$g_3(\hat{\mathbf{x}}) = -5.0397;$$

Obviously, we find the equality $g_2(\hat{\mathbf{x}}) = g_3(\hat{\mathbf{x}})$.

The inequality $g_1(\hat{\mathbf{x}}) > g_2(\hat{\mathbf{x}}) = g_3(\hat{\mathbf{x}})$ tells us that the second hyper-plane is **not** a decision boundary between class 2 and 3, as the probability for class 1 is greater than the probability for classes 2 and 3.

<u>Therefore, this is NOT a decision boundary, since $g_1(\hat{\mathbf{x}}) > g_j(\hat{\mathbf{x}}), j = 2,3$.</u>

Third decision boundary (between class 1 and class 3)**:**

$$g_1(\boldsymbol{x}) = g_3(\boldsymbol{x}) \Rightarrow \ln\frac{1}{4} = -2(x_1 + x_2) - 4 \; + \ln\frac{1}{2}$$

$$x_1 + x_2 = \frac{1}{2}\ln 2 - 2$$

$\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}$ is defined by equation $x_1 + x_2 == \frac{1}{2}\ln 2 - 2$

Substituting the above equation into the discriminant functions g1( ), g2 ( ) and g3 ( ) we obtain the value of these functions when g1( )= g3 ( ) :

$$g_1(\hat{\mathbf{x}}) = -1.3863;$$
$$g_2(\hat{\mathbf{x}}) = \text{-}8.6931;$$
$$g_3(\hat{\mathbf{x}}) = \text{-}1.3863;$$

Obviously, the value of g1 is constant, and we find the equality $g_1(\hat{\mathbf{x}}) = g_3(\hat{\mathbf{x}})$.

The inequality $g_2(\hat{\mathbf{x}}) < g_1(\hat{\mathbf{x}}) = g_3(\hat{\mathbf{x}})$ tell us that this is a decision boundary between class 1 and 3

Decision boundaries:

Between classes 1 and 2:

$$x_1 + x_2 = 2$$

Between classes 1 and 3:

$$x_1 + x_2 = \frac{1}{2}\ln 2 - 2 = -1.6534$$

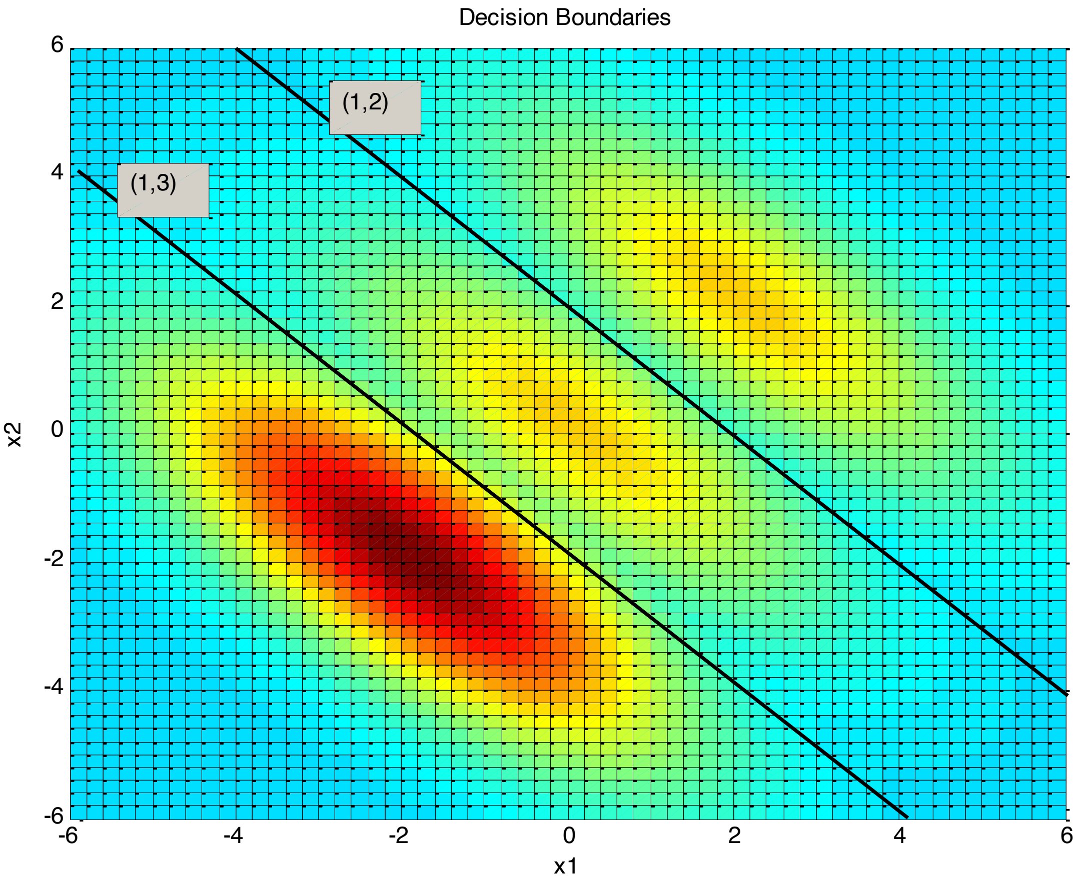


Fig.5 Probability densities and decision boundary. The decision boundary between class 1 and class 3 is identify by (1,3). The decision boundary between class 1 and class 2 is identify by (1,2).

It is easy to note that the point

$$x_c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

lies exactly on the hyper-plane (1,2), in fact this point can be assigned to class 1 or 2 (see the previous solution).

From the graphic above, it is easy to see that decision boundaries are not symmetric with respect to Gaussian centroids, due to the different class prior probabilities.

Exercise 4
Let us consider the following training examples for two classes ω_A and ω_B:

A	B
(-1, 0)^T	(5, 4)^T
(1, 0)^T	(3, 3)^T
(-0.1, 2)^T	(4, 7)^T
(0.3, -1.8)^T	(6.5, 2)^T

Class-conditional distributions are as follows:
$p(\mathbf{x}|\omega_A) = N(\mu_A, \Sigma_A)$; $p(\mathbf{x}|\omega_B) = N(\mu_B, \Sigma_B)$;

Note that the values of the parameters $\mu_A, \Sigma_A, \mu_B, \Sigma_B$ **are not given but they should be estimated.**

Question: find the optimal decision boundaries and classify the pattern $\hat{\mathbf{x}} =$ (2.2, 2.2)' under the following hypotheses:

1) $\Sigma_A = \Sigma_B = \sigma^2\mathbf{I}$ and $P(\omega_1) = P(\omega_2)$,
2) $\Sigma_A = \Sigma_B = \sigma^2\mathbf{I}$ and $P(\omega_1) = 2P(\omega_2)$
3) $\Sigma_A \neq \Sigma_B$ and $P(\omega_1) = 2P(\omega_2)$
4) $\Sigma_A = \Sigma_B$ **but with arbitrary form and** $P(\omega_1) = 2P(\omega_2)$

Note that changing the above assumptions, pattern classification can change as well.

SOLUTION

The general form of a multivariate normal p.d.f. is

$$p(x|\omega_i) = \frac{1}{(2\,\pi)^{\frac{d}{2}}\,|\Sigma_i|^{\frac{1}{2}}} \exp[-\frac{1}{2}\,(x-\mu_i)^T\,\Sigma_i^{-1}\,(x-\mu_i)]$$

We know from Chapter 4 that if we apply the logarithm function to the above expression, we obtain the following discriminant function $g_i(x)$:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\,\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

1) Under the hypothesis $\Sigma_A = \Sigma_B = \sigma^2\mathbf{I}$ and $P(\omega_1) = P(\omega_2)$, find the optimal discriminant function and classify the pattern $\hat{\mathbf{x}} =$ (2.2, 2.2)'

1.a classify the pattern
Under the above hypothesis, we know from Part 4 of the course that the discriminant functions are:

$$g_A(\mathbf{x}) = -\|\mathbf{x} - \boldsymbol{\mu}_A\|^2$$

$$g_B(\mathbf{x}) = -\|\mathbf{x} - \boldsymbol{\mu}_B\|^2$$

And the optimal decision boundary is:

$$g_A(\mathbf{x}) = g_B(\mathbf{x})$$

We can estimate the mean vectors (centroids of the two classes) using the equation:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k$$

$$\boldsymbol{\mu}_A = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix};\ \boldsymbol{\mu}_B = \begin{pmatrix} 4.625 \\ 4 \end{pmatrix}$$

Computing the values of the two discriminant functions above for the pattern $\hat{\mathbf{x}} =$ **(2.2, 2.2)'**:

(note that the discriminant functions are the squared distances of the pattern x from the class centroids:

$g_A(X)$=- 9.2450; $g_B(X)$=-9.1206;

Therefore, this pattern is assigned to class B.

1.b Find the decision boundary

The boundary is defined by the equation

$$||\boldsymbol{x} - \boldsymbol{\mu_A}||^2 = ||\boldsymbol{x} - \boldsymbol{\mu_B}||^2$$

where the generic 2-dimensional vector **x** is $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, that is,

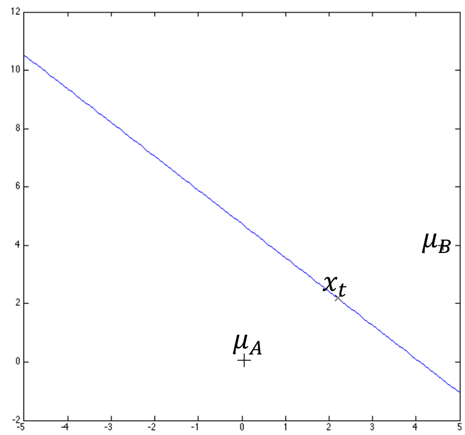
$$\left\|\begin{pmatrix} x_1 - \mu_{A1} \\ x_2 - \mu_{A2} \end{pmatrix}\right\|^2 = \left\|\begin{pmatrix} x_1 - \mu_{B1} \\ x_2 - \mu_{B2} \end{pmatrix}\right\|^2$$

$$(x_1 - \mu_{A1})^2 + (x_2 - \mu_{A2})^2 = (x_1 - \mu_{B1})^2 + (x_2 - \mu_{B2})^2$$

...

after a simple calculation, we obtain

$$9.15\ x_1 + 7.9\ x_2 = 37.3856$$



An alternative approach is

$$||\boldsymbol{x} - \boldsymbol{\mu_A}||^2 = ||\boldsymbol{x} - \boldsymbol{\mu_B}||^2$$

$$(\mathbf{x}-\boldsymbol{\mu}_A)^T(\mathbf{x}-\boldsymbol{\mu}_A)=(\mathbf{x}-\boldsymbol{\mu}_B)^T(\mathbf{x}-\boldsymbol{\mu}_B)$$

$$-\mathbf{x}^T\boldsymbol{\mu}_A-\boldsymbol{\mu}_A^T\mathbf{x}\ \ +\boldsymbol{\mu}_A^T\ \boldsymbol{\mu}_A=-\mathbf{x}^T\boldsymbol{\mu}_B-\boldsymbol{\mu}_B^T\mathbf{x}\ \ +\boldsymbol{\mu}_B^T\ \boldsymbol{\mu}_B$$

$$-2\boldsymbol{\mu}_A^T\ \mathbf{x}\ \ +\boldsymbol{\mu}_A^T\ \boldsymbol{\mu}_A=-2\boldsymbol{\mu}_B^T\ \mathbf{x}\ \ +\boldsymbol{\mu}_B^T\ \boldsymbol{\mu}_B$$

$9.15\ x_1 + 7.9\ x_2 = 37.3856$

The pattern (2.2, 2.2) is slightly above the boundary, and thus its is classified as belonging to class B (as discussed before).

2) Under the hypothesis $\Sigma_A = \Sigma_B = \sigma^2\mathbf{I}$ and $P(\omega_1) = 2P(\omega_2)$, find the optimal discriminant function and classify pattern $\hat{\mathbf{x}} =$ (2.2, 2.2)'

As in previous exercise, we can start our calculation from the general form of the discriminant function

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

by removing the terms common to both members:

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2}+\ln(P(\omega_i))$$

(the quadratic term is the same in each $g_i()$ and it can thus be ignored)

We need to estimate the term σ. The estimation of covariance matrices is

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{x}_k-\hat{\boldsymbol{\mu}}_i)(\mathbf{x}_k-\hat{\boldsymbol{\mu}}_i)^t$$

$$\hat{\Sigma}_A = \begin{pmatrix} 0.6967 & -0.25 \\ -0.25 & 2.41 \end{pmatrix}$$

$$\hat{\Sigma}_B = \begin{pmatrix} 2.2292 & -1.3333 \\ -1.3333 & 4.6667 \end{pmatrix}$$

But the hypothesis is $\Sigma_A = \Sigma_B = \sigma^2\mathbf{I}$

When all classes show the same covariance matrix, it can be proven that the maximum likelihood estimate for such a covariance matrix is (see slides of Part 4 of the course):

$$\widehat{\Sigma} = \sum_{i=1}^{c} \frac{n_i}{n} \widehat{\Sigma}_i$$

or, if we know the priors,

$$\widehat{\Sigma} = \sum_{i=1}^{c} P(\omega_i) \widehat{\Sigma}_i$$

In order to enforce compliance between the obtained data and the hypothesis $\Sigma_A = \Sigma_B = \sigma^2 \boldsymbol{I}$, we may impose that these matrices are diagonal and proportional to the identity matrix (only zero values outside the diagonal). Essentially, we have to ignore out-of-diagonal terms, and average terms on the main diagonal. This corresponds to computing the variance of each feature, and then average the corresponding values (separately for each class). Finally, the average variance is computed using the above equation (a weighted mean), namely, weighing the value of each average class variance with its prior probability.

$$\Sigma_A = \begin{pmatrix} 0.6967 & 0 \\ 0 & 2.41 \end{pmatrix};$$

$$\hat{\Sigma}_A = \begin{pmatrix} \frac{0.6967 + 2.41}{2} & 0 \\ 0 & \frac{0.6967 + 2.41}{2} \end{pmatrix} = 1.5533\, \boldsymbol{I}$$

$$\Sigma_B = \begin{pmatrix} 2.2292 & 0 \\ 0 & 4.6667 \end{pmatrix}$$

$$\hat{\Sigma}_B = \begin{pmatrix} \frac{2.2292 + 4.6667}{2} & 0 \\ 0 & \frac{2.2292 + 4.6667}{2} \end{pmatrix} = 3.4479\, \boldsymbol{I}$$

The above diagonal terms of the covariance matrix have been obtained by averaging the previous variance values.

Finally, $\widehat{\Sigma} = \sum_{i=1}^{c} P(\omega_i) \widehat{\Sigma}_i = P_A \widehat{\Sigma}_A + P_B \hat{\Sigma}_B = 2.1849\, \boldsymbol{I}$

so, $\sigma^2 = 2.1849$

2.a classify the pattern $\hat{\mathbf{x}} =$ (2.2, 2.2)’

$$g_A(x_t) = -\frac{\left||x_t - \mu_A|\right|^2}{2\,\sigma^2} + \log(P(\omega_A)) = -2.5212$$

$$g_B(x_t) = -\frac{||x_t - \mu_B||^2}{2\,\sigma^2} + \log(P(\omega_B)) = -3.1858$$

$g_B(\mathbf{x}) < g_A(\mathbf{x})$; **The pattern is assigned to class A**

Alternative approach

The discriminant functions can be expressed as

$g_i(x) = w_i^T \mathbf{x} + w_{i0}$, where

$$w_i = \frac{1}{\sigma^2}\mu_i; \;\; w_{i0} = -\frac{1}{2\sigma^2}\mu_i^T \mu_i + \log P(\omega_i)$$

(see chapter 4)

$$w_A = \begin{pmatrix} 0.0229 \\ 0.0229 \end{pmatrix}; w_{A0} = -0.4066$$

$$w_B = \begin{pmatrix} 2.1168 \\ 1.8308 \end{pmatrix}; w_{B0} = -9.6554$$

$$g_A(\mathbf{x}) = w_A^T \mathbf{x} + w_{A0} = -0.3059$$

$$g_B(\mathbf{x}) = w_B^T \mathbf{x} + w_{B0} = -0.9706$$

$g_B(\mathbf{x}) < g_A(\mathbf{x})$; **The pattern is assigned to class A**

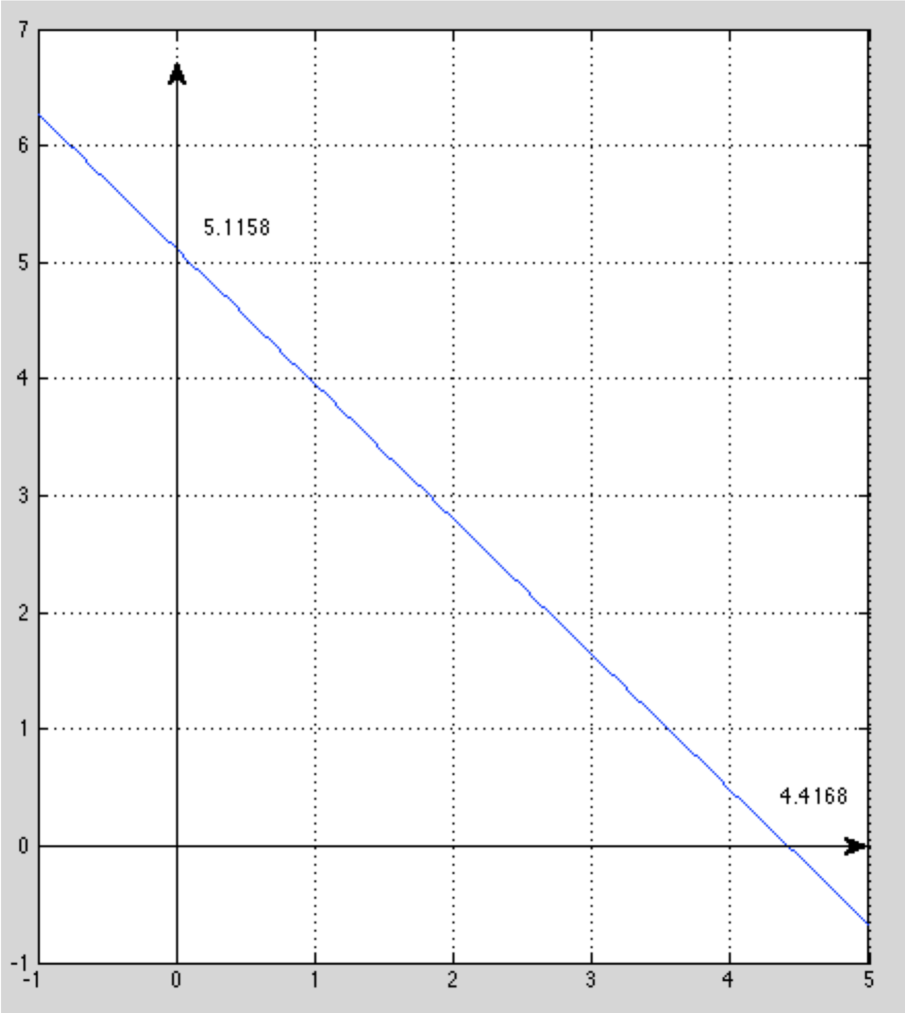
2.b find boundary

$$g_A(x) = -\frac{||x - \mu_A||^2}{2\,\sigma^2} + \log(P(\omega_A))$$

$$g_B(x) = -\frac{||x - \mu_B||^2}{2\,\sigma^2} + \log(P(\omega_B))$$

The optimal discriminant plane is the hyperplane $g_A = g_B$, that is,

$$2.094\, x_1 + 1.808\, x_2 = 9.2488$$



It is easy to see that (compare with point 1.b in the solution of this exercise) the boundary has been shifted towards the upper-right corner of the plot, due to the increase of the prior of the class A. Accordingly, the point (2.2, 2.2) is now below the decision boundary (thus in the decision region of class A).

(alternative approach)

The optimal discriminant plane is the hyperplane $g_A=g_B$, that is,

$$\boldsymbol{w}_A^T\,\boldsymbol{x} + w_{A0} = \boldsymbol{w}_B^T\,\boldsymbol{x} + w_{B0}$$

$$(\boldsymbol{w}_A^T - \boldsymbol{w}_B^T)\boldsymbol{x} + w_{A0} - w_{B0} = 0$$

The hyperplane can be written as

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0$$

that is,

$$2.094\;x_1 + 1.808\;x_2 = 9.2488$$

3) By assuming that $\Sigma_A \neq \Sigma_B$ e $P(\omega_1) = 2P(\omega_2)$, **find the optimal discriminant boundary and classify pattern** $\hat{\mathbf{x}} =$ (2.2, 2.2)'

We can estimate the covariance matrices (see previous exercise):

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^t$$

$$\Sigma_A = \begin{pmatrix} 0.6967 & -0.25 \\ -0.25 & 2.41 \end{pmatrix}$$

$$\Sigma_B = \begin{pmatrix} 2.2292 & -1.3333 \\ -1.3333 & 4.6667 \end{pmatrix}$$

$g_i(x)$ is a quadratic function which can be written as

$$g_i(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{W}_i\boldsymbol{x} + \boldsymbol{w}_i'\,\boldsymbol{x} + w_{i0}$$

where

$$\boldsymbol{W}_i = -\tfrac{1}{2}\,\boldsymbol{\Sigma}_i^{-1};\, w_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

Substituting:

$$\boldsymbol{W}_A = \begin{pmatrix} -0.7454 & -0.0773 \\ -0.0773 & -0.2155 \end{pmatrix}$$

wa = [0.0823, 0.0293]'

wa0 = -0.6484

$$\boldsymbol{W}_B = \begin{pmatrix} -0.2705 & -0.0773 \\ -0.0773 & -0.1292 \end{pmatrix}$$

wb =[3.1207 , 1.7487]'

wb0 = -12.8893

we can compute the discriminant functions and classify the pattern $\boldsymbol{x}_t = \begin{pmatrix} 2.2 \\ 2.2 \end{pmatrix}$

Discriminant function:

$g_a(\boldsymbol{x}) = -\,0.7454 x_1^2 \;-\; 0.2155\, x_2^2 - \;0.1547\, x_1 x_2 \;+\; 0.08227\, x_1 \;\;+\; 0.02928\, x_2 \;-\; 0.6484$

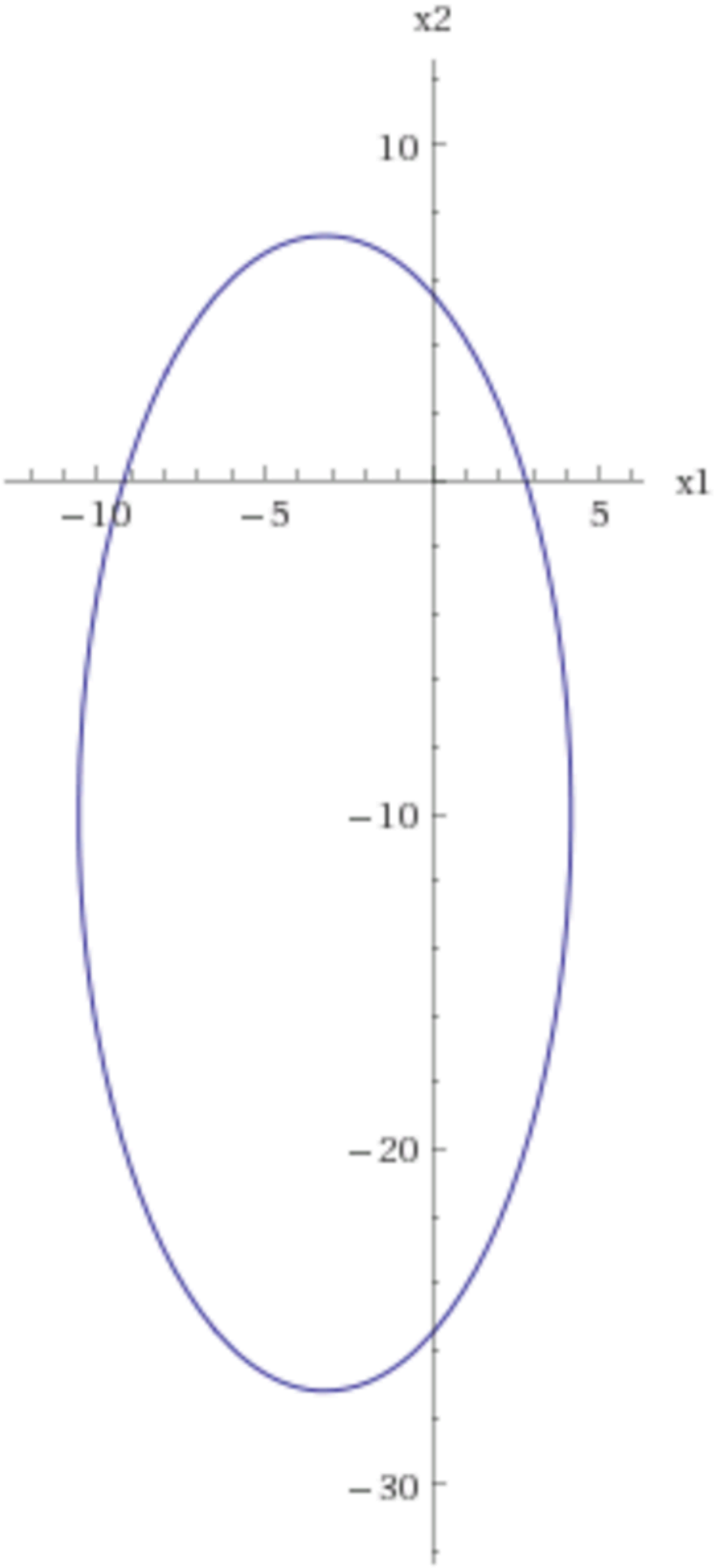
$g_b(\boldsymbol{x}) = -\,0.2705\, x_1^2 \;-\; 0.1546\, x_1 x_2 \;+\; 3.12\, x_1 \;-\; 0.1292\, x_2^2 \;+\; 1.749\, x_2 \;-\; 12.89$

Boundary:

As we know from SLIDES, chapt. 4, we obtain a quadratic discriminant function

$g_a(\boldsymbol{x}) - g_b(\boldsymbol{x}) = 0 \Rightarrow$

$\Rightarrow x_1^2 \;+\; 0.0001937\, x_1 x_2 \;+\; 6.397\, x_1 \;+\; 0.1816\, x_2^2 \;+\; 3.621\, x_2 \;-\; 25.78 = 0$



and finally

ga(**x_t**)= -5.825

gb(**x_t**)=-4 .8610

g_b(**x_t**)> g_A(**x_t**); **The pattern is assigned to class B**

4) $\Sigma_A = \Sigma_B$ **unknown and** $P(\omega_1) = 2P(\omega_2)$

Hint: we need to estimate the covariance matrix as a weighted mean of the two covariances and apply the Gaussian model which can satisfy the hypothesis $\Sigma_A = \Sigma_B$, as in Exercise 2.

Exercise 5

Let us consider a classification task with 3 data classes, a two-dimensional feature space, with Gaussian distributions for the 3 classes:

$$\boldsymbol{\mu}_1 = \begin{pmatrix}0\\0\end{pmatrix};\ \boldsymbol{\mu}_2 = \begin{pmatrix}0\\1\end{pmatrix};\ \boldsymbol{\mu}_3 = \begin{pmatrix}1\\1\end{pmatrix};\ for\ each\ class\ \boldsymbol{\Sigma}_i = 2\boldsymbol{I}$$

$$P(\omega_1) = P(\omega_2) = \frac{1}{4}$$

Question: Find the optimal decision boundaries and the related decision regions.

Solution

Given that $\boldsymbol{\Sigma_i = \sigma^2 I}$ we have $\boldsymbol{g_i(x)}$:

$$g_i(x) = -\frac{\left\|x-\mu_i\right\|^2}{2\,\sigma^2} + \log P_i$$

$$g_1(x) = -\tfrac{1}{4}(x_1^2 + x_2^2) + \log\tfrac{1}{4}$$

$$g_2(x) = -\tfrac{1}{4}(x_1^2 + x_2^2) - \tfrac{1}{4}(1 - 2\,x_2) + \log\tfrac{1}{4}$$

$$g_3(x) = -\tfrac{1}{4}(x_1^2 + x_2^2) - \tfrac{1}{4}(1 - 2\,x_1 + 1 - 2\,x_2) + \log\tfrac{1}{2}$$

Equal terms can be deleted, so obtaining:

$$g_1(x) = \log\tfrac{1}{4}$$

$$g_2(x) = -\tfrac{1}{4}(1 - 2\,x_2) + \log\tfrac{1}{4}$$

$$g_3(x) = -\tfrac{1}{4}(1 - 2\,x_1 + 1 - 2\,x_2) + \log\tfrac{1}{2}$$

Decision boundary between class 1 and class 2

$g_1(x) = g_2(x) \Longrightarrow x_2 = \frac{1}{2}$

For $\hat{\boldsymbol{x}} \equiv (x_2 = \frac{1}{2})$ we have

$g_1(\hat{\boldsymbol{x}}) = g_2(\hat{\boldsymbol{x}}) = \log\frac{1}{4}$;

$g_3(\hat{\boldsymbol{x}}) = -\frac{1}{4}(1 - 2\,x_1) + \log\frac{1}{2}$

$x_2 = \frac{1}{2}$ is the boundary if and only if $g_1(\hat{\boldsymbol{x}}) = g_2(\hat{\boldsymbol{x}}) > g_3(\hat{\boldsymbol{x}})$, therefore

Boundary between class 1 and 2

$$\begin{cases} \qquad x_2 = 1/2 \\ if\ x_1 < \frac{1}{2} - 2\log 2 \ \cong -0.8863 \end{cases}$$

Decision boundary between class 1 and class 3

$g_1(x) = g_3(x) \Longrightarrow x_1 + x_2 = 1 - 2\log 2$

For $\hat{\boldsymbol{x}} \equiv (x_1 + x_2 = 1 - 2\ \log 2)$ we have

$g_1(\hat{\boldsymbol{x}}) = g_3(\hat{\boldsymbol{x}}) = \log\frac{1}{4}$;

$g_2(\hat{\boldsymbol{x}}) = \frac{1}{2}x_2 - \frac{1}{4} + \log\frac{1}{4}$

$x_1 + x_2 = 1 - 2\ \log 2$ is the boundary only if $g_1(\hat{\boldsymbol{x}}) = g_3(\hat{\boldsymbol{x}}) > g_2(\hat{\boldsymbol{x}})$, therefore

Boundary between class 1 and class 3

$$\begin{cases} x_1 + x_2 = 1 - 2\log 2 \cong -0.3863 \\ \qquad if\ x_2 < \frac{1}{2} \end{cases}$$

Decision boundary between class 2 and classe 3

$g_2(x) = g_3(x) \Longrightarrow x_1 = \frac{1}{2} - 2\ \log 2$

For $\widehat{\boldsymbol{x}} \equiv (x_1 = \frac{1}{2} - 2\ \log 2\ \ )$ we have

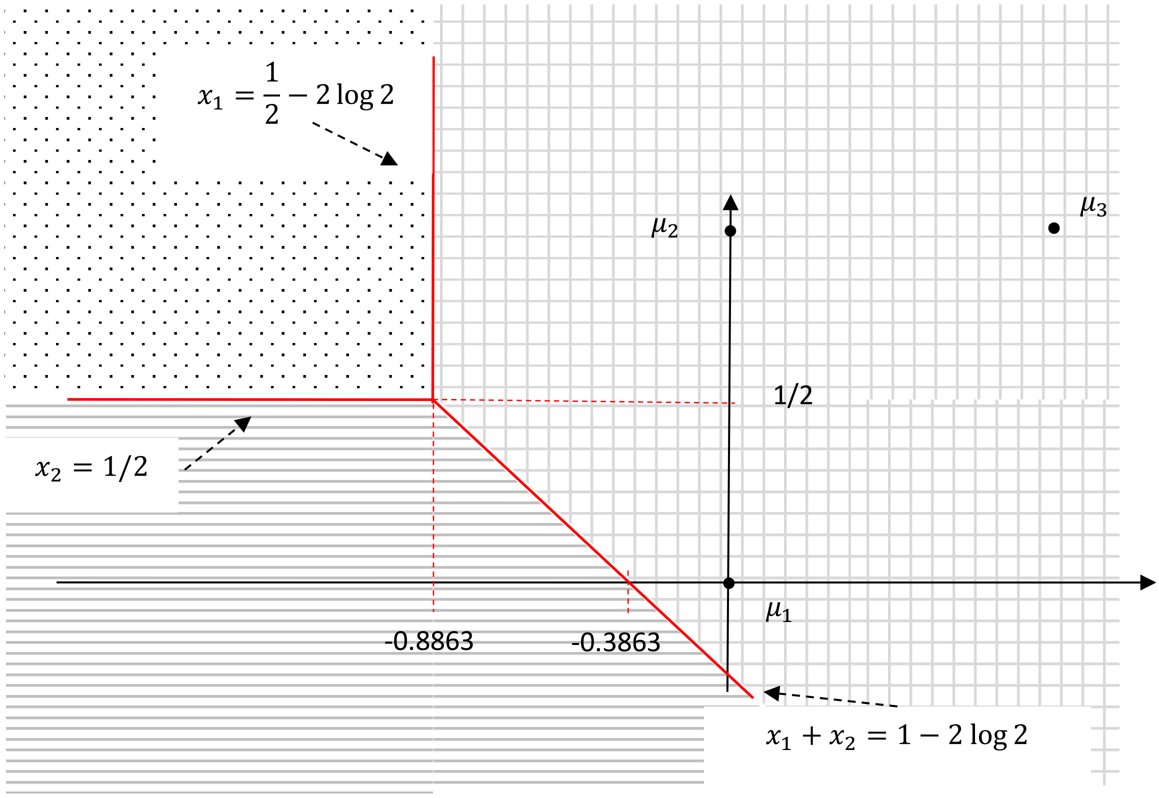
$g_2(\widehat{\boldsymbol{x}}) = g_3(\widehat{\boldsymbol{x}}) = -\frac{1}{4}(1 - 2\ x_2) + \log\frac{1}{4}$;

$g_1(\widehat{\boldsymbol{x}}) = \log\frac{1}{4}$

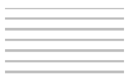
$-\frac{1}{4}(1 - 2\ x_2) + \log\frac{1}{4}$ is the boundary only if $g_2(\widehat{\boldsymbol{x}}) = g_3(\widehat{\boldsymbol{x}}) > g_1(\widehat{\boldsymbol{x}})$, therefore

Boundary class 2 and class 3

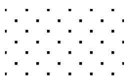
$$\begin{cases} x_1 = \frac{1}{2} - 2\log 2 \cong -0.8863 \\ \qquad\qquad se\ x_2 > \frac{1}{2} \end{cases}$$



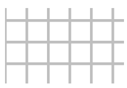
Legenda



class 1 decision region



class 2 decision region



class 3 decision region

Due to the difference of priors, both μ_1 and μ_2 lie in the decision region of class 3.