FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Machnine Learning (PDEEC0049 : 15-782PP)
Homework 1

António Damião das Neves Rodrigues (200400437 : 700098386)

February 20, 2014

1 PROBLEM 1

Following the definition of the conditional expectation of a random variable L, given the event $L > n$ with $P(L > n) > 0$, we get:

$$\mathbf{E}[L|L > n] = \sum_l l\,P(L|L > n) \tag{1.1}$$

Applying Bayes' theorem to 1.1 we get:

$$\mathbf{E}[L|L > n] = \sum_l l\,P(L|L > n) = \sum_l l\,\frac{P(L > n|L)P(L)}{P(L > n)}, \quad P(L > n) > 0 \tag{1.2}$$

The distribution of $P(L > n|L)$ can be easily determined by recognizing that if we are given a particular $L = l$, L is either larger than n or not larger than n, i.e.:

$$P(L > n|L = l) = \begin{cases} 1 & \text{if } l > n \\ 0 & \text{if } l \leq n \end{cases} \tag{1.3}$$

Applying 1.3 in the expression for $\mathbf{E}[L|L > n]$ (1.2), we get:

$$\mathbf{E}[L|L > n] = \sum_l l\,\frac{P(L > n|L)P(L)}{P(L > n)} = \sum_{l>n} l\,\frac{P(L)}{P(L > n)}, \quad P(L > n) > 0 \tag{1.4}$$

Considering $L = [l_{min}, l_{min+1}, ..., n-1, n, n+1, ..., l_{max-1}, l_{max}]$ where $l_{min} < l_{min+1} < ... < n-1 < n < n+1 < ... < l_{max-1} < l_{max}$, we can express $\mathbf{E}[L|L > n]$ in the form shown in 1.5:

$$
\begin{aligned}
\mathbf{E}[L|L>n] &= \sum_{l>n} l \frac{P(L)}{P(L>n)} \\
&= \frac{(n+1)P(L=n+1)+(n+2)P(L=(n+2))+\ldots+l_{max}P(L=l_{max})}{P(L=(n+1))+P(L=(n+2))+\ldots+P(L=l_{max})}, \quad P(L>n)>0
\end{aligned}
\tag{1.5}
$$

So $\mathbf{E}[L|L>n]$ produces a mean for values of L within the interval $]n, l_{max}]$. If $P(L>n)>0$, then $\mathbf{E}[L|L>n] \in]n, l_{max}]$.

The addition an element $l=(n-1)<n$ will contribute to lower the value of the average, i.e. $\mathbf{E}[L|L>n-1] < \mathbf{E}[L|L>n]$. The reduction of the value of the average occurs for every $l' < n$, such that $l' \geq l_{min}$, i.e. $\mathbf{E}\left[L|L>l'\right] < \mathbf{E}[L|L>n]$, with $l_{min} \leq l' < n$. We can consider $\mathbf{E}[L]$ as $\mathbf{E}\left[L|L>l'\right]$, when $l' \to l_{min}$.

Therefore $\mathbf{E}[L] = \mathbf{E}\left[L|L>l' \to l_{min}\right] < \mathbf{E}[L|L>n]$.

Looking at expression 1.5, if we consider the case in which $n \to l_{min}$, i.e. as $P(L>n) \to 1$ (and consequently $P(L \leq n) \to 0$), it can be seen that $\mathbf{E}[L|L>n] \to \mathbf{E}[L]$. Hence, here is why the inequality is strict for $P(L \leq n) > 0$.

2 Problem 2

2.1

Figure 2.1 depicts the scatter plots of the observations for each one of the sets X_1, X_2 and X_3 over the two-dimensional space (x_1, x_2).

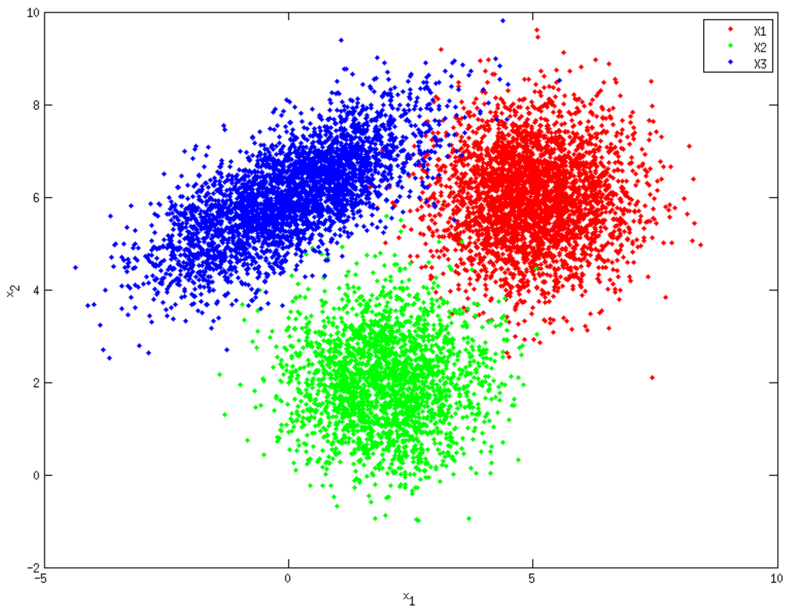

Figure 2.1: Graphical representation of the problem training set, consisting in three groups of observations X_1, X_2 and X_3.

2.2

The values of μ and $\sum$ (covariance matrix) for each one of the distributions of the observation sets X_1, X_2 and X_3 (obtained via MATLAB commands `mean` and `cov`) are shown in Table 2.1.

Observation Set	$\boldsymbol{\mu}$	$\sum$
X_1	$\begin{bmatrix} 5.0029 & 5.9756 \end{bmatrix}$	$\begin{bmatrix} 1.0161 & 0.0135 \\ 0.0135 & 1.0157 \end{bmatrix}$
X_2	$\begin{bmatrix} 1.9830 & 2.0029 \end{bmatrix}$	$\begin{bmatrix} 1.0183 & 0.0044 \\ 0.0044 & 1.0072 \end{bmatrix}$
X_3	$\begin{bmatrix} 0.0154 & 6.0173 \end{bmatrix}$	$\begin{bmatrix} 1.9229 & 0.9801 \\ 0.9801 & 0.9882 \end{bmatrix}$

Table 2.1: Values of μ and $\sum$ for each one of the distributions relative to the sets of observations X_1, X_2 and X_3.

By combined inspection of Figure 2.1 and Table 2.1, one can verify that the values of μ on the table are close to the centers of the respective 'swarms' of observations in the figure.

With μ and $\sum$ estimated from the data, we can now find the expressions for each of the models $N(\mathrm{x}|\mu_i, \sum_i)$:

$$N(\mathrm{x}|\mu_i, \sum{}_i) = \frac{e^{-\frac{1}{2}\left[(\mathrm{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\mathrm{x}-\mu_i)\right]}}{2\pi|\sum_i|^{\frac{1}{2}}} \tag{2.1}$$

These are models for the class-conditional distributions for each class C_i, i.e. each provides the value of the probability for **x**, given that the class is C_i. Therefore:

$$p(\mathrm{x}|C_i) = \frac{e^{-\frac{1}{2}\left[(\mathrm{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\mathrm{x}-\mu_i)\right]}}{2\pi|\sum_i|^{\frac{1}{2}}} \tag{2.2}$$

2.3

With the expressions for the class-conditional distributions, we can now apply Bayes' theorem (as given in (1.82) in [1]) to find the posterior probabilities $p(C_i|\mathrm{x})$:

$$p(C_i|\mathrm{x}) = \frac{p(\mathrm{x}|C_i)p(C_i)}{p(\mathrm{x})} \tag{2.3}$$

The prior probabilities $p(C_i)$, can inferred by considering the number of observations in each training set N_{X_i}, divided by the total number of points in the three training sets $N = N_{X_i} + N_{X_j} + N_{X_k}$:

$$p(C_i) = \frac{N_{X_i}}{N_{X_i} + N_{X_j} + N_{X_k}} = \frac{N_{X_i}}{N} \tag{2.4}$$

The boundary between classes C_i and C_k can be found by determining the common points between $p(C_i|\mathrm{x})$ and $p(C_k|\mathrm{x})$, i.e.:

$$\begin{aligned}
\frac{p(\mathrm{x}|C_i)p(C_i)}{p(\mathrm{x})} &= \frac{p(\mathrm{x}|C_j)p(C_j)}{p(\mathrm{x})} \\
\frac{p(C_i)e^{-\frac{1}{2}\left[(\mathrm{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\mathrm{x}-\mu_i)\right]}}{2\pi|\sum_i|^{\frac{1}{2}}} &= \frac{p(C_j)e^{-\frac{1}{2}\left[(\mathrm{x}-\mu_j)^{\mathrm{T}}\Sigma_j^{-1}(\mathrm{x}-\mu_j)\right]}}{2\pi|\sum_j|^{\frac{1}{2}}} \\
\frac{e^{-\frac{1}{2}\left[(\mathrm{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\mathrm{x}-\mu_i)\right]}}{e^{-\frac{1}{2}\left[(\mathrm{x}-\mu_j)^{\mathrm{T}}\Sigma_j^{-1}(\mathrm{x}-\mu_j)\right]}} &= \frac{p(C_j)\,|\sum_i|^{\frac{1}{2}}}{p(C_i)|\sum_j|^{\frac{1}{2}}} \\
e^{-\frac{1}{2}\left[(\mathrm{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\mathrm{x}-\mu_i)-(\mathrm{x}-\mu_j)^{\mathrm{T}}\Sigma_j^{-1}(\mathrm{x}-\mu_j)\right]} &= e^{\ln\left(\frac{p(C_j)|\Sigma_i|^{\frac{1}{2}}}{p(C_i)|\Sigma_j|^{\frac{1}{2}}}\right)}
\end{aligned} \tag{2.5}$$

Both sides of equation 2.5 are equal when the exponents are the same. By evaluating the exponents only, we get equation 2.6:

$$-\frac{1}{2}\left[\left(\mathrm{x}-\mu_i\right)^{\mathrm{T}}\sum_i^{-1}\left(\mathrm{x}-\mu_i\right)-\left(\mathrm{x}-\mu_j\right)^{\mathrm{T}}\sum_j^{-1}\left(\mathrm{x}-\mu_j\right)\right]-\ln\left(\frac{p(C_j)\,|\Sigma_i|^{\frac{1}{2}}}{p(C_i)\,|\Sigma_j|^{\frac{1}{2}}}\right)=0 \tag{2.6}$$

For the boundary $\mathbf{u_{ij}}$ between classes $\boldsymbol{C_i}$ and $\boldsymbol{C_j}$, we therefore have expression 2.6, which is a curve on two-dimensional space (x_1, x_2). Boundaries $\mathbf{u_{12}}$, $\mathbf{u_{13}}$ and $\mathbf{u_{23}}$ are given by expressions 2.7, 2.8 and 2.9 respectively:

$$-\frac{1}{2}\left[\left(\mathrm{x}-\mu_1\right)^{\mathrm{T}}\sum_1^{-1}\left(\mathrm{x}-\mu_1\right)-\left(\mathrm{x}-\mu_2\right)^{\mathrm{T}}\sum_2^{-1}\left(\mathrm{x}-\mu_2\right)\right]-\ln\left(0.7\frac{|\Sigma_1|^{\frac{1}{2}}}{|\Sigma_2|^{\frac{1}{2}}}\right)=0 \tag{2.7}$$

$$-\frac{1}{2}\left[\left(\mathrm{x}-\mu_1\right)^{\mathrm{T}}\sum_1^{-1}\left(\mathrm{x}-\mu_1\right)-\left(\mathrm{x}-\mu_3\right)^{\mathrm{T}}\sum_3^{-1}\left(\mathrm{x}-\mu_3\right)\right]-\ln\left(\frac{|\Sigma_1|^{\frac{1}{2}}}{|\Sigma_3|^{\frac{1}{2}}}\right)=0 \tag{2.8}$$

$$-\frac{1}{2}\left[\left(\mathrm{x}-\mu_2\right)^{\mathrm{T}}\sum_2^{-1}\left(\mathrm{x}-\mu_2\right)-\left(\mathrm{x}-\mu_3\right)^{\mathrm{T}}\sum_3^{-1}\left(\mathrm{x}-\mu_3\right)\right]-\ln\left(1.4\frac{|\Sigma_2|^{\frac{1}{2}}}{|\Sigma_3|^{\frac{1}{2}}}\right)=0 \tag{2.9}$$

Let $f_{ij}(x)$ be the left side of equations 2.7, 2.8 and 2.9. For every class pair evaluation, given an observation x, we should decide for class C_i if $f_{ij}(x) \geq 0$ or class C_j if $f_{ij}(x) < 0$. Note that each $f_{ij}(x)$ determines if an observation x belongs to class C_i, considering the border with class C_j. In order to completely determine the class for x, we must evaluate the remaining border via $f_{ik}(x)$ or $f_{jk}(x)$.

One strategy to follow in order to assign an observation x to a class C_1, C_2 or C_3 can be as shown below. Note that for $N = 3$ classes the proposed nested set of conditional statements is feasible, however such an approach might get too complex for a larger N.

for all observation x **do**
    if $f_{12}(x) \geq 0$ **then**
        if $f_{31}(x) \geq 0$ **then**
            assign observation x to class C_3
        else
            assign observation x to class C_1
        end if
    else if $f_{23}(x) \geq 0$ **then**
        assign observation x to class C_2
    else
        assign observation x to class C_3
    end if
end for

2.4

Listings 1 and 2 provide the MATLAB code used for classifying the observation set X_x. The results are graphically shown in Figure 2.2 (same color code as in Figure 2.1). The class boundaries obtained in 2.3 are also plotted. (using MATLAB command `ezplot()`).

Listing 1: MATLAB code for 2.4.

```
clear

% load the problem's data
load('.../trainingset.mat');

% total number of observations from the training sets
N = length(X1(:,1)) + length(X2(:,1)) + length(X3(:,1));

% calculate the parameters for each one of the distributions Xk, to be
% used in the discriminant functions dscrmnt(x, ...)

% mean
U1 = mean(X1);

% covariance matrix
Cov1 = cov(X1);

% inverse of the covariance matrix (let's calculate it beforehand and not
% every time dsrmnt(x, ...) is ran)
iCov1 = inv(Cov1);

% the factor which affects the Gaussian distributions of Xk, including
% the prior probability for each class, pri
pr1 = length(X1(:,1))/(N);
k1 = pr1/(2*pi*sqrt(det(Cov1)));

U2 = mean(X2);
Cov2 = cov(X2);
iCov2 = inv(Cov2);
pr2 = length(X2(:,1))/(N);
k2 = pr2/(2*pi*sqrt(det(Cov2)));

U3 = mean(X3);
Cov3 = cov(X3);
iCov3 = inv(Cov3);
pr3 = length(X3(:,1))/(N);
k3 = pr3/(2*pi*sqrt(det(Cov3)));

% for convenience, create arrays to accomodate the each observation
% belonging to class Ck
C1 = [];
C2 = [];
C3 = [];


```

```
figure;
hold on;

[r,c] = size(XX);

% for each observation x in XX, determine its class Ck following the
% heuristics defined in 2.3 b)
for i=1:r

    x = XX(i,:);

    % descriminant function template for C1 and C2
    if dscrmnt(x,U1,U2,iCov1,iCov2,log(k2/k1)) == 1

        % use the descriminant function template for C3 and C1, to
        % evaluate the boundary u13
        if dscrmnt(x,U3,U1,iCov3,iCov1,log(k1/k3)) == 1

            % have found an observation which falls into C3
            % add an entry to the appropriate class array
            C3 = [C3; x];

        else

            % x indeed belongs to C1
            C1 = [C1; x];

        end

    % descriminant function template for C2 and C3
    elseif dscrmnt(x,U2,U3,iCov2,iCov3,log(k3/k2)) == 1

        % x indeed belongs to C1
        C2 = [C2; x];

    else

        % x belongs to C3
        C3 = [C3; x];

    end

end

% print the axis labels
xlabel('x_1');
ylabel('x_2');

% plot the observations in Xx, the color indicates the class to which an
% observation was assigned (same color code as in 2.1)
plot(C1(:,1), C1(:,2), 'r.');
plot(C2(:,1), C2(:,2), 'g.');
plot(C3(:,1), C3(:,2), 'b.');

```

```
% let's plot the class boundaries
syms x1 x2
u12=-0.5*(([x1 x2] - U1)*iCov1*([x1; x2]-U1') - ([x1 x2] - ...
    U2)*iCov2*([x1; x2]-U2')) - log(k2/k1);
u13=-0.5*(([x1 x2] - U1)*iCov1*([x1; x2]-U1') - ([x1 x2] - ...
    U3)*iCov3*([x1; x2]-U3')) - log(k3/k1);
u23=-0.5*(([x1 x2] - U2)*iCov2*([x1; x2]-U2') - ([x1 x2] - ...
    U3)*iCov3*([x1; x2]-U3')) - log(k3/k2);

pu12 = ezplot(u12,[2,8,-2,6]);
h = get(gca,'children');
set(h(1),'linestyle','-.','color','b','linewidth',1);

pu13 = ezplot(u13,[2,5,4,12]);
h = get(gca,'children');
set(h(1),'linestyle','-.','color','g','linewidth',1);

pu23 = ezplot(u23,[-3,2,-2,6]);
h = get(gca,'children');
set(h(1),'linestyle','-.','color','r','linewidth',1);

% color code for the observation-class assignment
legend('C_1','C_2','C_3','u12','u13','u23');
```

Listing 2: MATLAB code for 2.4.

```
% template for the discriminant function between class a vs. class b.
% returns 1 if x is classified as belonging to class 'a', 0 if 'b'.

% classes a and b which are discriminated by dscrmnt() are determined by
% setting the appropriate input parameters Uk, iCovk and C.

% e.g. the discriminant function between classes 1 and 2 is
% dscrmnt(x,U1,U2,iCov1,iCov2,C12).
function [dscrmnt] = dscrmnt(x,Ua,Ub,iCova,iCovb,C)

if (-0.5*((x-Ua)*iCova*(x-Ua)' - (x-Ub)*iCovb*(x-Ub)') - C) ≥ 0

    dscrmnt = 1;

else

    dscrmnt = 0;

end
```

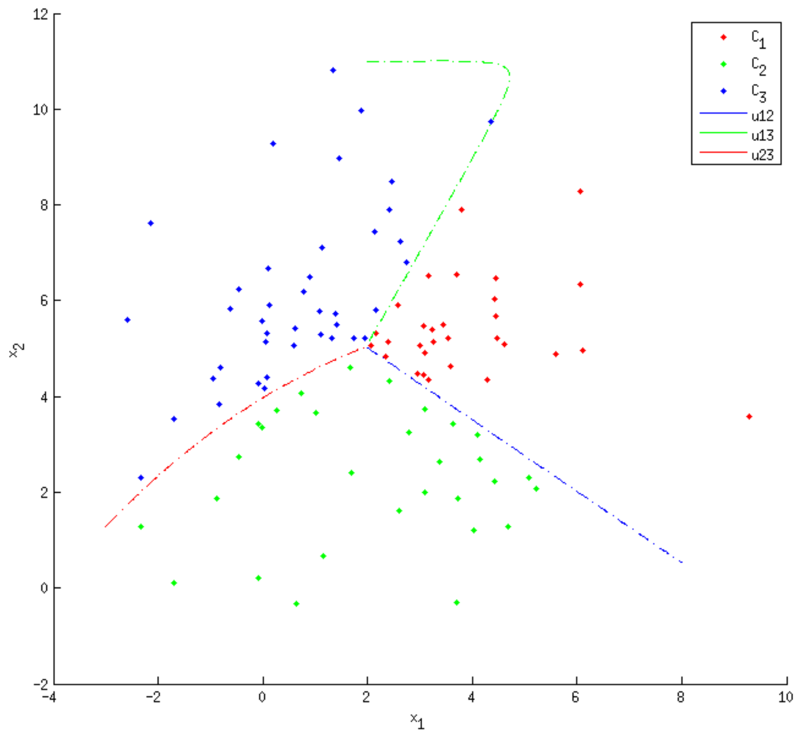

Figure 2.2: Graphical representation of the classification of the observations in set X_x. The class boundaries obtained in 2.3 are also plotted.

2.5

The exercise asks for way to calculate the probability of a selected class given an observation, i.e. the posterior probability $p(C_i|x)$. This probability can be found by using Bayes' theorem (as given in (1.82) in [1]):

$$p(C_i|\mathbf{x}) = \frac{p(\mathbf{x}|C_i)p(C_i)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|C_i)p(C_i)}{\sum_i p(\mathbf{x}|C_i)p(C_i)} \tag{2.10}$$

The class-conditional probabilities for each class C_i are obtained via the estimations performed in Section 2.2.

The prior probabilities $p(C_i)$, can be inferred as indicated in Section 2.3, i.e. by considering the number of observations in each training set N_{X_i}, divided by the total number of points in the three training sets $N = N_{X_i} + N_{X_j} + N_{X_k}$:

$$p(C_i) = \frac{N_{X_i}}{N_{X_i} + N_{X_j} + N_{X_k}} = \frac{N_{X_i}}{N} \tag{2.11}$$

With the class-conditional and prior probabilities determined, Bayes' theorem should be used to find $p(C_k|x)$.

3 Problem 3

This is a classification problem for two classes C_A and C_B, and a random variable x , with part of the inference stage parameters already determined: the class-conditional densities $p(x|C_A)$ and $p(x|C_B)$, as well as the prior class probabilities $p(C_A)$ and $p(C_B)$. Therefore, approach (a) given in Section 1.5.4 of [1] is appropriate to tackle this problem.

3.1

With part of the inference stage out of the way, one should now use Bayes' theorem to determine the posterior class probabilities $p(C_A|x)$ and $p(C_B|x)$, which will allow us to go ahead into the decision phase. Applying Bayes' theorem (as given in (1.82) in [1]), we get:

$$p(C_A|x) = \frac{p(x|C_A)p(C_A)}{p(x)} = \frac{0.4e^{-x}}{p(x)}, \quad x \geq 0 \tag{3.1}$$

$$p(C_B|x) = \frac{p(x|C_B)p(C_B)}{p(x)} = \frac{\frac{0.6}{\sqrt{2\pi}}e^{-(x-2)^2}}{p(x)}, \quad x \geq 0 \tag{3.2}$$

From the combination of the sum and product rules of probability (see (1.10) and (1.11) in [1]) and using the given expressions for the conditional and prior probabilities, we get the following expression for $p(x)$:

$$p(x) = \sum_k p(x|C_k)p(C_k) = 0.4e^{-x} + \frac{0.6}{\sqrt{2\pi}}e^{-(x-2)^2}, \quad x \geq 0 \tag{3.3}$$

Applying 3.3 in equations 3.1 and 3.2, we get the final expressions for $p(C_A|x)$ and $p(C_B|x)$:

$$p(C_A|x) = \frac{0.4e^{-x}}{0.4e^{-x} + \frac{0.6}{\sqrt{2\pi}}e^{-(x-2)^2}}, \quad x \geq 0 \tag{3.4}$$

$$p(C_B|x) = \frac{\frac{0.6}{\sqrt{2\pi}}e^{-(x-2)^2}}{0.4e^{-x} + \frac{0.6}{\sqrt{2\pi}}e^{-(x-2)^2}}, \quad x \geq 0 \tag{3.5}$$

With the posterior probabilities $p(C_A|x)$ and $p(C_B|x)$, we can now determine the decision boundaries $\hat{x}$ and consequently the decision regions $\mathscr{R}$ which minimize the misclassification rate. In this case, the same reasoning as that used in Section 1.5.1 of [1] can be applied, i.e. in order to minimize the misclassification rate, each value of x should be assigned to the class for which the probability $p(C_k|x)$ is the largest.

We should then find the values of x for which $p(C_k|x)$ intersect in order to determine the N decision boundaries $\hat{x}_n$ and then evaluate which posterior probability $p(C_k|x)$ has the highest value along each one of the $N+1$ resulting decision regions $\mathscr{R}_n$:

$$
\begin{aligned}
p(C_A|x) &= p(C_B|x) \\
0.4e^{-x} &= \frac{0.6}{\sqrt{2\pi}} e^{-(x-2)^2} \\
\frac{e^{-x}}{e^{-(x-2)^2}} &= e^{\ln\left(\frac{3}{2\sqrt{2\pi}}\right)} \\
e^{x^2-5x+4} &= e^{\ln\left(\frac{3}{2\sqrt{2\pi}}\right)}, \quad x \geq 0
\end{aligned}
\tag{3.6}
$$

Both sides of equation 3.6 are equal when the exponents are the same. By evaluating the exponents only, we get equation 3.7:

$$
x^2 - 5x + \left(4 - \ln\left(\frac{3}{2\sqrt{2\pi}}\right)\right) = 0, \quad x \geq 0 \tag{3.7}
$$

Solving 3.7 one gets the following two values of x for the boundaries, $\hat{x}_1$ and $\hat{x}_2$:

$$
\hat{x}_1 = \frac{5 - \sqrt{25 - 4\left(4 - \ln\left(\frac{3}{2\sqrt{2\pi}}\right)\right)}}{2} \approx 1.1822 \qquad \hat{x}_2 = \frac{5 + \sqrt{25 - 4\left(4 - \ln\left(\frac{3}{2\sqrt{2\pi}}\right)\right)}}{2} \approx 3.8178
$$

We therefore have three regions: $\mathscr{R}_1$, $\mathscr{R}_2$ and $\mathscr{R}_3$, separated by boundaries $\hat{x}_1$ and $\hat{x}_2$ respectively. Evaluating the values of $p(C_k|x)$ in the vicinities of both $\hat{x}_1$ and $\hat{x}_2$, one gets the decision regions given in Table 3.1.

Region	**Class**	**Interval**[1]
$\mathscr{R}_1$	C_A	$0 \leq x < \hat{x}_1$
$\mathscr{R}_2$	C_B	$\hat{x}_1 \leq x < \hat{x}_2$
$\mathscr{R}_3$	C_A	$\hat{x}_2 \leq x < +\infty$

[1] The inclusions/exclusions of points $x = \hat{x}_1$ and $x = \hat{x}_2$ into/from each one of the regions was arbitrary.

Table 3.1: Decision regions of the Bayes classifier.

Figure 3.1 provides a graphical representation of the proposed solution, including decision boundaries and probability density functions.

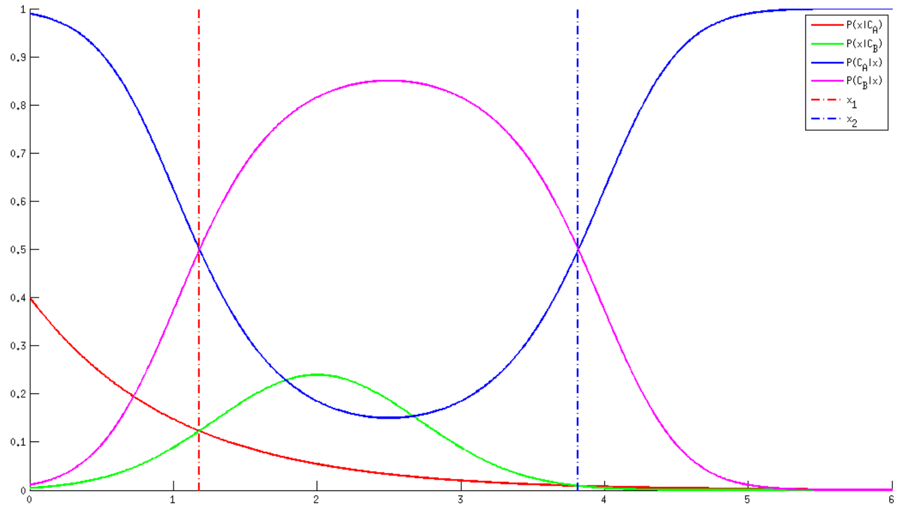

Figure 3.1: Graphical representation of the proposed solution, including decision boundaries and probability density functions.

3.2

Considering $x = 1$, since $x \leq \hat{x}_1$, it lies in decision region $\mathscr{R}_1$, therefore the predicted class for this observation is C_A. The probability of error can be determined by the value of the posterior probability for class C_B given $x = 1$, $p(C_B|x = 1)$. Applying $x = 1$ in 3.5, we get $p(C_B|x = 1) \approx 0.3744$.

3.3

We define the loss matrix L as shown below (each element L_{kj} is the penalty value for assigning an observation x to class C_j although it belongs to C_k):

$$L = \begin{bmatrix} L_{AA} & L_{AB} \\ L_{BA} & L_{BB} \end{bmatrix} = \begin{bmatrix} 0 & 1.2 \\ 0.8 & 0 \end{bmatrix} \tag{3.8}$$

Intuitively, due to the higher penalty value L_{AB}, one expects the values for the new boundaries $\hat{x}$' to shift towards the middle of the region $\mathscr{R}_2$ calculated in Section 3.1, i.e. to see the size of the region associated with class C_B reduced.

The loss one gets for deciding for one of the classes C_k, $L(C_k)$, is given by expressions 3.9 and 3.10.

$$L(C_A) = L_{AA}p(C_A|x) + L_{BA}p(C_B|x), \quad x \in \mathscr{R}_A \tag{3.9}$$

$$L(C_B) = L_{AB}\,p(C_A|x) + L_{BB}\,p(C_B|x), \quad x \in \mathscr{R}_B \tag{3.10}$$

For a given value of x, one should decide for the class C_k which presents the lowest loss value $L(C_k)$. Following the same reasoning as in Section 3.1, we should then find the values of x for which $L(C_k)$ intersect in order to determine the N new decision boundaries $\hat{x}'_n$ and then evaluate which loss $L(C_k)$ has the lowest value along each one of the $N+1$ resulting decision regions $\mathscr{R}'{}_n$:

$$\begin{aligned}
L(C_A) &= L(C_B)\\
L_{AA}\,p(C_A|x) + L_{AB}\,p(C_A|x) &= L_{BA}\,p(C_B|x) + L_{BB}\,p(C_B|x)\\
0 + \frac{0.48e^{-x}}{p(x)} &= \frac{\frac{0.48}{\sqrt{2\pi}}e^{-(x-2)^2}}{p(x)} + 0\\
0.48e^{-x} &= \frac{0.48}{\sqrt{2\pi}}e^{-(x-2)^2}\\
e^{x^2-5x+4} &= e^{\ln\left(\frac{1}{\sqrt{2\pi}}\right)}
\end{aligned} \tag{3.11}$$

By evaluating the exponents in 3.11, we get equation 3.12:

$$x^2 - 5x + \left(4 - \ln\left(\frac{1}{\sqrt{2\pi}}\right)\right) = 0, \quad x \geq 0 \tag{3.12}$$

Solving 3.12 one gets the following two values of x for the boundaries, $\hat{x}'_1$ and $\hat{x}'_2$:

$$\hat{x}'_1 = \frac{5 - \sqrt{25 - 4\left(4 - \ln\left(\frac{1}{\sqrt{2\pi}}\right)\right)}}{2} \approx 1.3463 \qquad \hat{x}'_2 = \frac{5 + \sqrt{25 - 4\left(4 - \ln\left(\frac{1}{\sqrt{2\pi}}\right)\right)}}{2} \approx 3.6537$$

We therefore have three new decision regions: $\mathscr{R}'_1$, $\mathscr{R}'_2$ and $\mathscr{R}'_3$, separated by boundaries $\hat{x}'_1$ and $\hat{x}'_2$ respectively. As expected, region $\mathscr{R}'_2$ is smaller than $\mathscr{R}_2$ determined in Section 3.1. The decision regions are summarized in Table 3.2.

Region	**Class**	**Interval**[1]
$\mathscr{R}'_1$	C_A	$0 \leq x < \hat{x}'_1$
$\mathscr{R}'_2$	C_B	$\hat{x}'_1 \leq x < \hat{x}'_2$
$\mathscr{R}'_3$	C_A	$\hat{x}'_2 \leq x < +\infty$

[1] The inclusions/exclusions of points $x = \hat{x}'_1$ and $x = \hat{x}'_2$ into/from each one of the regions was arbitrary.

Table 3.2: New decision regions of the Bayes classifier, taking the costs in matrix L (see 3.8) into account.

REFERENCES

[1] C.M. Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006.