Machine Learning Course
Example of intermediate assessment – year 2018

Students should do all the exercises to get the maximum score.
If you solve all the three exercises correctly, you get 33 points.
Please, justify carefully each answer.

Name: ……………………………. Surname: ……………………………. Student ID: …………

Exercise 1 (13 points)

Let's consider a 2-class problem in a one-dimensional feature space.
The class-conditional probability densities of the 2 classes are:

$p(x|\omega_1) = N(x;\ \mu_1 = -1, \sigma_1 = 1)$
$p(x|\omega_2) = N(x;\ \mu_2 = +1, \sigma_2 = 1)$

Recall that $N(x;\ \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, and that $\int_{-\infty}^{x} N(x;\ \mu, \sigma)\, dx = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)$.

Assume that the prior probabilities of the two data classes are $P(\omega_1) = 0.5P(\omega_2)$, and that the cost of errors of class ω_1 is double than those of class ω_2, that is, $\lambda_{21} = 2\lambda_{12}$ (and $\lambda_{11} = \lambda_{22} = 0$).

- Compute the minimum-risk decision regions and, separately, the two errors (one per class).
- Compare these errors with the corresponding errors achieved using the Bayesian optimal threshold.
- Justify the results. *Is the Bayesian error higher/lower than the error obtained with the minimum-risk approach? Why?*

Exercise 2 (9 points)

Let us assume that $P(\omega_1) = P(\omega_2)$, while $p(x|\omega_i)$ is defined as in Exercise 1. We want to minimize the classification error (not the risk!) using the rejection option with the Chow's rule.
Let us assume that the rejection region is defined as [-0.5, 0.5].

- Compute the rejection threshold T on the posterior probabilities.
- Compute the fraction of rejected samples of class 2.
- Compute the two errors, separately for each class.

Exercise 3 (11 points)

Let us consider a 3-class problem in R^2 (two-dimensional feature space), where the likelihood of each class is Gaussian and given as $p(x|\omega_i) = N(\boldsymbol{\mu_i}, \Sigma_i)$, with

$\Sigma_i = \sigma^2\mathbf{I}$; $\mu 1 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$; $\mu 2 = \begin{pmatrix} +1 \\ -1 \end{pmatrix}$; $\mu 3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$;

and prior probabilities $P_1 = P_2 = P_3$.

- Compute the decision boundaries and plot them.

Solutions

Exercise 1

The a-priori probabilities are $P(\omega_1) = 1/3$; $\mathrm{P}(\omega_2) = 2/3$

The minimum-risk decision rule amounts to deciding for class 1 if

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} > \left(\frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}\right)\frac{P(\omega_2)}{P(\omega_1)}$$

Noting that the standard deviation is equal to 1 for both classes, and costs of correct decisions are 0, we can simplify the above expression as:

$$\frac{\exp\left(-\frac{|x-\mu_1|^2}{2\sigma^2}\right)}{\exp\left(-\frac{|x-\mu_2|^2}{2\sigma^2}\right)} > \frac{\lambda_{12}}{\lambda_{21}}\frac{P(\omega_2)}{P(\omega_1)}$$

In this case, $\frac{\lambda_{12}}{\lambda_{21}}\frac{P(\omega_2)}{P(\omega_1)} = 1$, but let's set it generically to *k*.

By recalling that $\sigma^2 = 1$ for both classes, and taking the logarithm on both sides we simply get

$$-|x-\mu_1|^2 + |x-\mu_2|^2 > 2\ln(k)$$
$$-x^2 + 2x\mu_1 - \mu_1^2 + x^2 - 2x\mu_2 + \mu_2^2 > 2\ln(k)$$

Since $\mu_1^2 = \mu_2^2 = 1$, we finally have: $x(\mu_1 - \mu_2) > \ln(k)$

Now, setting k=1, and substituting $\mu_1 = -1$ and $\mu_2 = 1$, we find that $-2x < 0$, that is, when $x < 0$, the optimal decision is to assign x to class 1 (otherwise, assign x to class 2).

Therefore, the minimum-risk decision regions are:

$R_1 \equiv (-\infty, 0)$; $R_2 \equiv [0, +\infty)$ (0 is assigned here to class 2 but assigning it to class 1 is also valid).

Let's compute now the errors separately per class.

Error probability for patterns belonging to class 1:

$$E_1 = P_1 \int_{R_2} p(x|\omega_1) = P_1 \int_0^{+\infty} N(x;\ \mu_1 = -1, \sigma_1 = 1)\,dx = \frac{1}{3}\left(\frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{0-\mu_1}{\sqrt{2}\sigma_1}\right)\right) = \frac{1}{6} - \frac{1}{6}\operatorname{erf}\left(\frac{\sqrt{2}}{2}\right)$$
$$= 0.0529$$

Error probability for patterns belonging to class 2:

$$E_2 = P_2 \int_{R_1} p(x|\omega_2) = P_2 \int_{-\infty}^{0} N(x;\ \mu_2 = 1, \sigma_2 = 1)\,dx = \frac{2}{3}\left(\frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{0-\mu_2}{\sqrt{2}\sigma_2}\right)\right) = \frac{1}{3} + \frac{1}{3}\operatorname{erf}\left(-\frac{\sqrt{2}}{2}\right)$$
$$= 0.1058$$

To compute the Bayesian decision regions, it suffices to recall that the optimal decision is to assign the pattern to class 1 if $x(\mu_1 - \mu_2) > \ln(k)$ where *k here* does not include the costs, namely, $k = \frac{P(\omega_2)}{P(\omega_1)} = 2$

Accordingly, we obtain

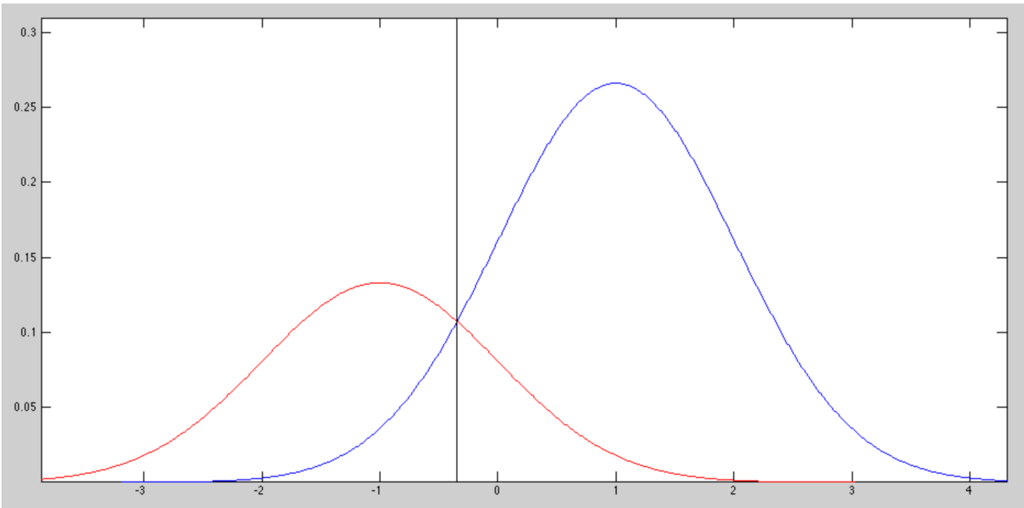
$$x(\mu_1 - \mu_2) > \ln(2)$$
$$-2x > \ln(2)$$
$$x < -\frac{1}{2}\ln(2) => x^* = -0.3466$$

Therefore, the Bayesian decision regions are:
$R_1 \equiv (-\infty, x^*)$; $R_2 \equiv [x^*, +\infty)$ (as before, $x = x^*$ is assigned here to class 2 but assigning it to class 1 is also valid). The corresponding Bayesian errors are obtained using the previous formulas, but replacing 0 with the value $x^* = -0.3466$

Error probability for patterns belonging to class 1:
$$E_1 = P_1 \int_{R_2} p(x|\omega_1) = \frac{1}{6} - \frac{1}{6}\mathrm{erf}\left(\frac{x^* - \mu_1}{\sqrt{2}\sigma_1}\right) = 0.0856$$

Error probability for patterns belonging to class 2:
$$E_2 = P_2 \int_{R_1} p(x|\omega_2) = \frac{2}{3}\left(\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{x^* - \mu_2}{\sqrt{2}\sigma_2}\right)\right) = \frac{1}{3} + \frac{1}{3}\mathrm{erf}\left(\frac{x^* - \mu_2}{\sqrt{2}}\right) = 0.0594$$

Compare the errors obtained by the minimum-risk approach with the errors obtained with the Bayesian threshold. Justify the results.
When we set $\lambda_{21} = 2\lambda_{12}$, the threshold is shifted from the Bayesian threshold -0.3466 to 0, to enlarge the decision region of class 1 and thus reduce the number of patterns of class 1 misclassified as class 2. Clearly, the total error increases from the Bayesian error of 0.145 to 0.165.
The plot shows the probability distributions $P(\omega)p(x|\omega)$ for the two classes and the Bayesian threshold.



Matlab code to generate the plot
```
plot(x, 1/3*normpdf(x,-1,1), 'r'); hold on
plot(x, 2/3*normpdf(x,1,1), 'b');
stem(-0.3466,1, 'k')
```

Exercise 2
The rejection threshold T is computed by equating the posterior probability at the boundary of the rejection region.

Let's consider class 1:

$p(\omega_1|x = -0.5) = \; p(x = -0.5 \,|\omega_1)p(\omega_1)/p(x = -0.5)$

where $p(\omega_1) = 0.5$, $p(x = -0.5 \,|\omega_1) = 0.3521$, and
$p(x = -0.5) = p(\omega_1)p(x = -0.5|\omega_1) + p(\omega_2)p(x = -0.5|\omega_2) = 0.5(0.3521 + 0.1295) = 0.2408$

Thus, $T = p(\omega_1|x = -0.5) = 0.7311$

For symmetry, the same holds for class 2, namely $T = p(\omega_2|x = 0.5) = 0.7311$

Fraction of rejected patterns of class 2:
$E_2 = P_2 \int_{R_r} p(x|\omega_2) = \; P_2 \int_{-0.5}^{+0.5} N(x;\; \mu_2 = 1, \sigma_2 = 1)\; dx$
The Gaussian integral can be computed by integrating the two tails and then subtracting them from 1

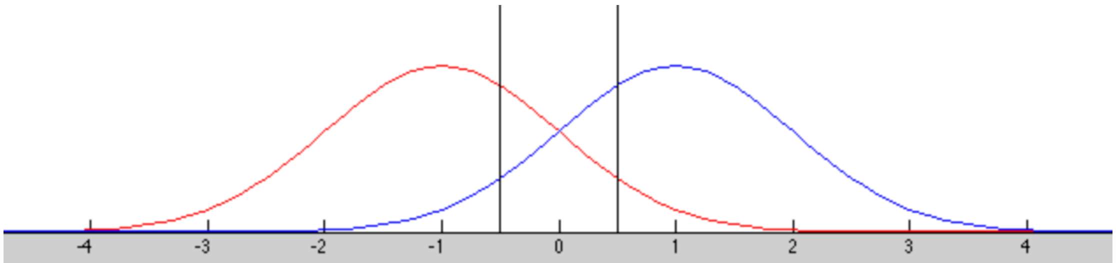
$$\int_{-0.5}^{+0.5} N(x;\; \mu_2, \sigma_1)\; dx$$

$$= 1 - \int_{-\infty}^{-0.5} N(x;\; \mu_2, \sigma_2) dx - \int_{0.5}^{+\infty} N(x;\; \mu_2, \sigma_2) dx = 1 - \left(\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{-0.5-1}{\sqrt{2}}\right)\right)$$
$$- \left(\frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{0.5-1}{\sqrt{2}}\right)\right) = \frac{1}{2}\mathrm{erf}\left(\frac{1.5}{\sqrt{2}}\right) - \frac{1}{2}\mathrm{erf}\left(\frac{0.5}{\sqrt{2}}\right) = 0.2417$$

Thus, $E_2 = P_2 * 0.2417 = 0.1209$ *(for symmetry, it is the same for class 1)*

Error probability for patterns belonging to class 1:
$E_1 = P_1 \int_{R_2} p(x|\omega_1) = \; P_1 \int_{0.5}^{+\infty} N(x;\; \mu_1 = -1, \sigma_1 = 1)\; dx = \frac{1}{2}\left(\frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{0.5+1}{\sqrt{2}}\right)\right) = \frac{1}{4} - \frac{1}{4}\mathrm{erf}\left(\frac{3\sqrt{2}}{4}\right)$
$= 0.0334$
(for symmetry, it is the same for class 2)

Error probability for patterns belonging to class 2:
$E_2 = P_2 \int_{R_1} p(x|\omega_2) = \; P_2 \int_{-\infty}^{-0.5} N(x;\; \mu_2 = 1, \sigma_1 = 1)\; dx = \frac{1}{4} + \frac{1}{4}\mathrm{erf}\left(-\frac{3\sqrt{2}}{4}\right)$
$= 0.0334$



Exercise 3

The generalized discriminant function for Gaussian distributions is:

$$g(x) = -\frac{1}{2}x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu + \ln p(\omega) - \frac{1}{2}\ln|\Sigma|$$

In this case, the covariance matrix is isotropic, and equal for all classes. Even the priors are the same. Thus, the above expression can be simplified as: $g(x) = \mu^T x - \frac{1}{2}\mu^T\mu$

Notably, the second term is also equal for all classes ($\mu^T\mu = 2$ for all classes), and thus this term can also be removed from the discriminant function. Therefore, we obtain: $g(x) = \mu^T x$.
Accordingly, for each class we have
$g_1(x) = \mu_1^T x = -x_1 - x_2$; $g_2(x) = \mu_2^T x = +x_1 - x_2$; $g_3(x) = \mu_3^T x = +x_1 + x_2$

Let us now compute the class boundaries between each pair of classes.

Class boundary between class 1 and class 2. We start by finding x^* for which $g_1(x^*) = g_2(x^*)$:
$(\mu_1 - \mu_2)^T x = 0$, which implies $x_1 = 0$

This boundary is aligned with the y-axis (i.e., the x_2 values) and holds only for the subset of points for which it holds that
$g_1(x^*) = g_2(x^*) > g_3(x^*)$
$-x_2 > x_2$ which implies $x_2 < 0$
This boundary is thus aligned with and active for the non-positive part of the y-axis.

Class boundary between class 1 and class 3. Let us find the points x^* for which $g_1(x^*) = g_3(x^*)$:
$(\mu_1 - \mu_3)^T x = 0$, which implies $x_1 = -x_2$

The boundary holds only for the subset of points for which it holds that
$g_1(x^*) = g_3(x^*) > g_2(x^*)$, i.e., $0 > x_1 - x_2$
Now, if we substitute $x_1 = -x_2$ in the above inequality, we have $x_2 > 0$.
Alternatively, one may substitute $x_2 = -x_1$ in the above inequality and obtain $x_1 < 0$.
These conditions are clearly equivalent, as together with the boundary equation $x_1 = -x_2$ both identify the top-left quadrant of the cartesian space (where the boundary is active).

Class boundary between class 2 and class 3. Let us find the points x^* for which $g_2(x^*) = g_3(x^*)$:
$(\mu_2 - \mu_3)^T x = 0$, which implies $x_2 = 0$

$g_2(x) = \mu_2^T x = +x_1 - x_2$; $g_3(x) = \mu_3^T x = +x_1 + x_2$

This boundary is aligned with the x-axis (i.e., the x_1 values) and holds only for the subset of points for which it holds that $g_2(x^*) = g_3(x^*) > g_1(x^*)$. This implies that $x_1 > 0$
This boundary is thus aligned with and active for the positive part of the x-axis

This plot shows the class boundaries for the three Gaussian classes with $\sigma = 0.3$