Machine Learning Course
First intermediate assessment – April 15, 2019

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 (10 points)
Let's consider a 2-class problem in a one-dimensional feature space.
The class-conditional probability densities of the 2 classes are:
$p(x|\omega_1) = N(x;\ \mu_1 = -2, \sigma_1 = 2)$
$p(x|\omega_2) = N(x;\ \mu_2 = +1, \sigma_2 = 1)$
Recall that $N(x;\ \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, and that $\int_{-\infty}^{x} N(x;\ \mu, \sigma)\, dx = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)$.
Assume that the prior probabilities of the two data classes are $P(\omega_1) = P(\omega_2)$, 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 $\lambda_{11} = \lambda_{22} = 0$).

- (5 points) Compute the minimum-risk decision regions, and the Bayesian decision regions.
- (3 points) Plot the joint distributions $P_k\, p(x|\omega_k)$ for k=1,2 on the one-dimensional feature space, along with the minimum-risk and the Bayesian decision regions.
- (2 points) In the plot, highlight the area(s) corresponding to the additional error incurred when using the minimum-risk decision regions rather than the Bayesian ones. One of the two classes exhibits a higher error. *Which one? Why?*

Exercise 2 (12 points)
Let us assume that the probability distributions are 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 [-2, 6].

- Compute the rejection threshold T on the posterior probabilities, for class 1 and class 2 on both the bounds of the reject region.
- Why the values of T computed for the two classes are different? Plot (approximately) the posterior distributions w.r.t. x to find the answer.
- Compute the fraction of rejected samples of class 2.

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
$\Sigma_i = \sigma^2\mathbf{I}$; $\mu 1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$; $\mu 2 = \begin{pmatrix} +1 \\ -1 \end{pmatrix}$; $\mu 3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$;
and prior probabilities $P_1 = P_2 = P_3$.

- Compute the decision boundaries and plot them.

Solutions

Exercise 1

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

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:

$$\frac{\exp\left(-\frac{|x-\mu_1|^2}{2\sigma^2}\right)}{\exp\left(-\frac{|x-\mu_2|^2}{2\sigma^2}\right)} > \frac{\sigma_1}{\sigma_2}\frac{\lambda_{12}}{\lambda_{21}}\frac{P(\omega_2)}{P(\omega_1)} = k$$

In this case, $\mathrm{k} = 1$, but let's set it generically to k. Taking the log on both sides one yields:

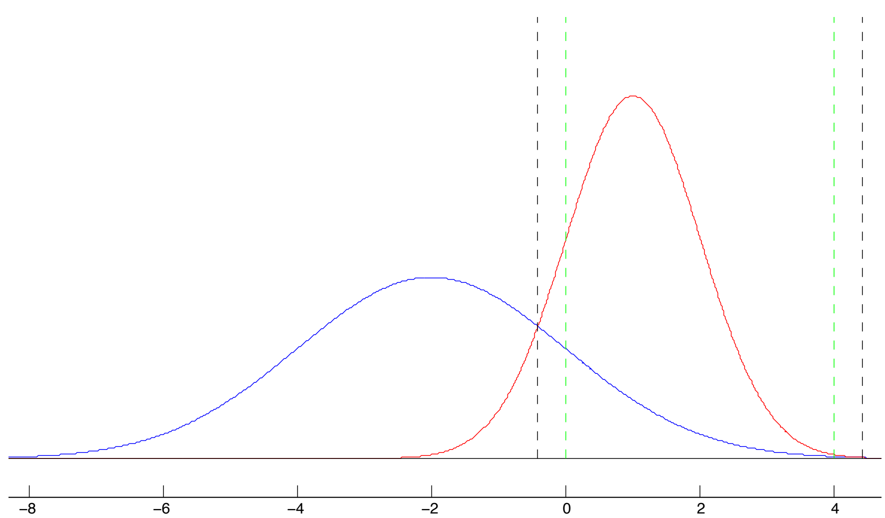
$$-\frac{|x-\mu_1|^2}{2\sigma_1^2} + \frac{|x-\mu_2|^2}{2\sigma_2^2} = \ln(k)$$

Replacing the values of the parameters and of k, and solving w.r.t. x, we get two values:
$x_{1/2} = 2 \pm \sqrt{4}$, i.e., $x_1 = 0$ and $x_2 = 4$.
Therefore, the minimum-risk decision regions are: $R_1 \equiv (-\infty, x_1) \cup (x_2, +\infty);\ R_2 \equiv [x_1, x_2]$.

To compute the Bayesian decision regions, it suffices to set k as k=ln(2) and solve again the given second-order equation obtained from the likelihood ratio, obtaining:

$x_{1/2} = 2 \pm \sqrt{4 + \frac{8}{3}\ln(2)}$, i.e., $x_1 = -0.41$ and $x_2 = 4.41$.
Therefore, the Bayesian decision regions are: $R_1 \equiv (-\infty, x_1) \cup (x_2, +\infty);\ R_2 \equiv [x_1, x_2]$

The plot shows the Bayesian decision boundaries in black, and the minimum-risk ones in green. It is clear from the plot that, when considering the minimum-risk decision regions, the error on class 2 increases, as R_2 is reduced (due to a larger cost of misclassifying class 1 as 2).



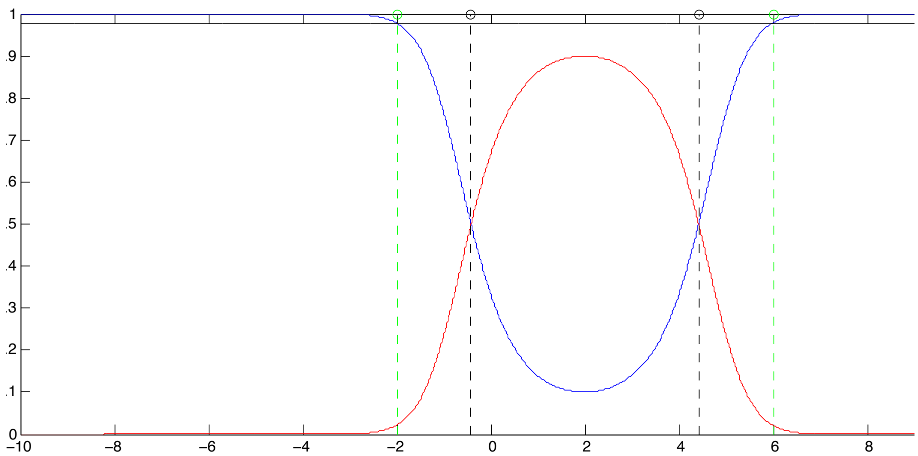
Exercise 2

The rejection threshold T is computed by equating the posterior probability at the boundary of the rejection region. Let's consider class 1:

$T = p(\omega_1|x = -2) = p(x = -2\,|\omega_1)\frac{p(\omega_1)}{p(x=-2)} = 0.9783$. The same value is found for $p(\omega_1|x = 6)$.

For class 1, computing the posterior in -2 and 6, one instead yields: 0.0217.

What is exactly happening here? The plot below explains the situation:



It is easy to see that class 2 never exceeds the reject threshold T, and thus no sample will be ever assigned to class 2 in this case.

Fraction of rejected patterns of class 2:

$E_2 = P_2 \int_{R_r} p(x|\omega_2) = P_2 \int_{-2}^{+6} N(x;\ \mu_2 = 1, \sigma_2 = 1)\, dx$

The Gaussian integral can be computed as

$$\int_{-2}^{+6} N(x;\ \mu_2, \sigma_2)dx = \int_{-\infty}^{6} N(x;\ \mu_2, \sigma_2)dx - \int_{-\infty}^{-2} N(x;\ \mu_2, \sigma_2)dx =$$

$$= \frac{1}{2}\mathrm{erf}\left(\frac{6-1}{\sqrt{2}}\right) - \frac{1}{2}\mathrm{erf}\left(\frac{-2-1}{\sqrt{2}}\right) = 0.9986$$

Thus, the result is $E_2 = 0.499$

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: $g(x) = \mu^T x - \frac{1}{2}\mu^T\mu$

Accordingly, for each class we have

$g_1(x) = -x_1 - \frac{1}{2}$; $g_2(x) = x_1 - x_2 - 1$; $g_3(x) = \mu_3^T x = x_1 + x_2 - 1$

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

This boundary holds only for the subset of points for which it holds that

$g_1(x^*) = g_2(x^*) > g_3(x^*)$ which implies $x_2 < 0$

This boundary is thus active for the negative 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 $x_2 = -2x_1 + \frac{1}{2}$

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., $x_2 > 0$.

This boundary is thus active for the positive part of the y-axis.

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 $x_2 = 0$

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 $x_1 > \frac{1}{4}$

This boundary is thus aligned with the x-axis, and active for $x_1 > \frac{1}{4}$

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