Exercise 1

In the two-category case, under the Bayes decision rule the conditional error is given by $P(error|x) = min[P(\omega_1|x), P(\omega_2|x)]$. Even if the posterior densities are continuous, this form of the conditional error virtually always leads to a discontinuous integrand when calculating the full error by $P(error) = \int P(error|x)p(x)dx$.

a) Show that for arbitrary densities, we can replace the first equation by $P(error|x) = 2P(\omega_1|x)P(\omega_2|x)$ in the integral and get an upper bound on the full error.

b) Show that if we use $P(error|x) = \alpha P(\omega_1|x)P(\omega_2|x)$ for $\alpha < 2$, then we are not guaranteed that the integral gives an upper bound on the error.

Exercise 2

A mobile robot is required to navigate towards a goal position while avoiding possible collisions with moving obstacles. The robot can perform only two actions: *maneuvre* (α_1) to surround the obstacles or *stop* (α_2) at the current position if the obstacles are too close. The *maneuvre* action modifies the robot's desired orientation θ_d according to:

$$\theta_d = \begin{cases} \theta_r + \theta_{av} & \text{if } d < d_{safe} \\ \theta_r & \text{otherwise} \end{cases} \tag{1}$$

where θ_r is the measured orientation of the robot, θ_{av} is the rotation needed to surround the obstacles and d is the distance between the robot and the obstacles. All the obstacles at a distance $d > d_{safe}$ do not affect the robot's path.

At each time instant, the state of path the robot is executing can be: *obstacle* (ω_1) or *no_obstacle* (ω_2). In a first stage, a human is remotely guiding the robot towards the path. From the collected observations, the robot has learned that $p(\omega_1) = 0.7$ and $p(\omega_2) = 0.3$. Assume as costs $c_{11} = 0$, $c_{12} = c_{21} = 5$ and $c_{22} = 10$, summarized in the following table:

	ω_1	ω_2
α_1	0	5
α_2	5	10

The robot is equipped with a range sensor that generates a 3D point cloud. Points at a distance $d < d_{safe} = 1m$ from the robot (x_1) are considered as obstacles, points at a distance $d \geq d_{safe} = 1m$ (x_2) are considered as free-space (*no_obstacle*). Due to the noisy data d cannot be accurately estimated, suppose: $p(x_1|\omega_1) = 0.8$ and $p(x_2|\omega_2) = 0.7$.

a) Comment the choice of the costs c_{11} and c_{22}.

b) Using the Bayes risk criterion determine which is the best action considering the observations from the range sensor.

Exercise 3

Given two-dimensional data and two categories with parameters μ_1 and Σ_1 and μ_2 and Σ_2, calculate the decision boundaries.

It is given that $\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

Exercise 4

As a result of a dichotomous classification of diseases, the patient might have heart disease (1) or not (0). We want to study the effect of smoking on the heart disease problem. The additional independent variables which enter the problem are race, sex and other three health conditions $x_1 = CAT$, $x_2 = EGG$ and $x_3 = AGE$. If the age of a person is equal to a, then $x_3 = AGE$ is computed as $x_3 = \frac{1}{3}\log a$. The logistic function that relates the variables to the disease is $p(x) = \frac{1}{(1+e^{-X})}$ where $x = a_0 + a_1x_1 + ...a_3x_3$. Information is gathered for 700 white males over 10 years. We suppose that after learning the model, the following results are derived:

a) $a_0 = 4$, $a_1 = 0.7$, $a_2 = 0.03$, $a_3 = 0.4$. What is the probability with which a 40-years old person with CAT=1 and EGG=0 is at heart disease risk.

b) In statistics the quantity $\frac{p(X)}{1-p(X)}$ is called *odds*, and it reflects the likelihood that a particular event will take place. The natural logarithm of the odds is called $logit\,(p(x)) = \ln\frac{p(X)}{1-p(X)}$ represents the odds for a person developing the disease with independent variable x. Compute the odds for the above condition (Ex. 3-a).

c) If all of x variables are zero, what does $logit\,(p(x))$ show?

Exercise 5

The following table shows 4 training samples from a survey. Two attributes have been selected to classify data samples as good or bad.

x_1	x_2	$y(classification)$
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

An incoming sample is the $(x_1 = 3, x_2 = 7)$. Classify the sample by using k-nearest neighbor method.