Lehrstuhl für STEUERUNGS- UND REGELUNGSTECHNIK Technische Universität München Prof. Dongheui Lee	**MACHINE LEARNING IN ROBOTICS** Exercises : Classifier and Clustering

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.

Solution Exercise 1

a) The conditional probability of error in a two category case is given by:

$$p(error|x) = \begin{cases} p(\omega_1|x) & if\ we\ decide\ \omega_2 \\ p(\omega_2|x) & if\ we\ decide\ \omega_1 \end{cases} \tag{1}$$

Following the Bayes Decision Rule in this case (choose ω_1 if $p(\omega_1|x) > p(\omega_2|x)$) the conditional probability of error is $p(error|x) = min[p(\omega_1|x), p(\omega_2|x))]$, and it is optimal, as it ensures that the total probability of error $\int_{-\infty}^{\infty} p(error|x)p(x)dx$ will be minimal by minimizing the $p(error|x)$ function.

We want to find an upper bound on the total probability of error by proving that

$$2p(\omega_1|x)p(\omega_2|x) > p(error|x) \tag{2}$$

in any case.

As we are in a two category case, we can divide the whole range in two regions, one where $p(\omega_1|x) < p(\omega_2|x)$ and another one where $p(\omega_2|x) < p(\omega_1|x)$. Based on this division, we could split the error as follows:

$$\int_{-\infty}^{\infty} p(error|x)p(x)dx = \int\limits_{p(\omega_1|x)<p(\omega_2|x)} p(\omega_1|x)p(x)dx \ + \int\limits_{p(\omega_2|x)<p(\omega_1|x)} p(\omega_2|x)p(x)dx \tag{3}$$

We just need to proof now that the given bound is valid for both regions. In that case, it will be valid for the whole integral.

Let us consider that we are in the first region, and, therefore, $\underline{p(\omega_1|x) < p(\omega_2|x)}$. In this region our error is always $p(error|x) = p(\omega_1|x)$. If we check that our bound is always bigger than this conditional error we will have an upper bound on the integral of this region. So, let us proof, $\underline{2p(\omega_1|x)p(\omega_2|x) > p(\omega_1|x)}$.

As $\sum\limits_{i=1}^{n.\ classes} p(\omega_i|x) = 1$, we can write our bound as:

$$2p(\omega_1|x)p(\omega_2|x) = 2p(\omega_1|x)(1 - p(\omega_1|x)) = 2(p(\omega_1|x) - p(\omega_1|x)^2) \tag{4}$$

As we are in the first region $p(\omega_1|x) < p(\omega_2|x)$ and with only 2 classifiers $p(\omega_1|x) < 0.5$, so:

$$p(\omega_1|x)^2 < \frac{p(\omega_1|x)}{2} \tag{5}$$

Merging equations 4 and 5, we could finally write:

$$2p(\omega_1|x)p(\omega_2|x) = 2(p(\omega_1|x) - p(\omega_1|x)^2) \ > \ 2(p(\omega_1|x) - \frac{p(\omega_1|x)}{2}) = p(\omega_1|x) \tag{6}$$

With eq.6 we have proven that our upper bound is valid for the first region.
With the same procedure we could prove the same for region 2 where $p(\omega_2|x) < p(\omega_1|x)$ and with it, the initial condition at eq.2 would be proven for the whole integral eq.3.

b) Starting at eq.6 and changing the 2 with α, we have:

$$\alpha(p(\omega_1|x) - p(\omega_1|x)^2) \ > \ \alpha(p(\omega_1|x) - \frac{p(\omega_1|x)}{2}) = \alpha(\frac{p(\omega_1|x)}{2}) \geq p(\omega_1|x) \tag{7}$$

To satisfy the two final steps of eq.7, $\alpha \geq 2$.

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:

$$\theta_d = \begin{cases} \theta_r + \theta_{av} & \text{if } d < d_{safe} \\ \theta_r & \text{otherwise} \end{cases} \tag{8}$$

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.

Solution Exercise 2

a) $c_{11} = 0 \longrightarrow$ do not penalize the *maneuvre* action when there are obstacles.

$c_{22} = 10 \longrightarrow$ strongly penalize the *stop* action when there are no obstacles.

b) Given $p(x_1|\omega_1) = 0.8$ and $p(x_2|\omega_2) = 0.7$ we can compute $p(x_2|\omega_1) = 1 - p(x_1|\omega_1) = 0.2$ and $p(x_1|\omega_2) = 1 - p(x_2|\omega_2) = 0.3$. The Bayes risk is defined as:

$$R(\alpha_i|x) = \sum_{j=1}^{2} c_{ij} p(\omega_j|x), \quad i = 1, \ldots, 2$$

hence, the only unknowns are the posteriors $p(\omega_j|x)$. First, calculate $p(x_1)$ and $p(x_2)$:

$$\begin{aligned} p(x_1) &= p(x_1|\omega_1)p(\omega_1) + p(x_1|\omega_2)p(\omega_2) = 0.8 * 0.7 + 0.3 * 0.3 = 0.65 \\ p(x_2) &= 1 - p(x_1) = 1 - 0.65 = 0.35 \end{aligned}$$

Then, using the Bayes rule:

$$\begin{aligned} R(\alpha_1|x_1) &= p(\omega_1|x_1)c_{11} + p(\omega_2|x_1)c_{12} = 0 + p(\omega_2|x_1) * 5 \\ &= \frac{p(x_1|\omega_2)p(\omega_2)}{p(x_1)} * 5 = \frac{0.3 * 0.3}{0.65} * 5 = 0.69 \\ R(\alpha_2|x_1) &= p(\omega_1|x_1)c_{21} + p(\omega_2|x_1)c_{22} = p(\omega_1|x_1) * 5 + p(\omega_2|x_1) * 10 \\ &= \frac{p(x_1|\omega_1)p(\omega_1)}{p(x_1)} * 5 + \frac{p(x_1|\omega_2)p(\omega_2)}{p(x_1)} * 10 = \frac{0.8 * 0.7}{0.65} * 5 + \frac{0.3 * 0.3}{0.65} * 10 = 4.3 + 1.39 = 5.69 \\ R(\alpha_1|x_2) &= p(\omega_1|x_2)c_{11} + p(\omega_2|x_2)c_{12} = 0 + p(\omega_2|x_2) * 5 \\ &= \frac{p(x_2|\omega_2)p(\omega_2)}{p(x_2)} * 5 = \frac{0.7 * 0.3}{0.35} * 5 = 3 \\ R(\alpha_2|x_2) &= p(\omega_1|x_2)c_{21} + p(\omega_2|x_2)c_{22} = p(\omega_1|x_2) * 5 + p(\omega_2|x_2) * 10 \\ &= \frac{p(x_2|\omega_1)p(\omega_1)}{p(x_2)} * 5 + \frac{p(x_2|\omega_2)p(\omega_2)}{p(x_2)} * 10 = \frac{0.2 * 0.7}{0.35} * 5 + \frac{0.7 * 0.3}{0.35} * 10 = 2 + 6 = 8 \end{aligned}$$

From the Bayes risk is obvious that the robot always chooses the *maneuvre* action.

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

Solution Exercise 3

For the discriminant function, we use the formula

$$g_i(x) = -\tfrac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \tfrac{1}{2}\ln|\Sigma_i| + \ln p(w_i)$$

Since $p(w_i)$ is the same for both categories, we can ignore this term for the comparison.

Class w_1:

$$g_1(x) = -\tfrac{1}{2}\begin{bmatrix} x_1-3 & x_2-6 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x_1-3 \\ x_2-6 \end{bmatrix} - \tfrac{1}{2}\ln 1.$$

Class w_2:

$$g_2(x) = -\tfrac{1}{2}\begin{bmatrix} x_1-3 & x_2+2 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x_1-3 \\ x_2+2 \end{bmatrix} - \tfrac{1}{2}\ln 4.$$

To estimate the decision boundary, we solve the equation

$$g_1(x) = g_2(x)$$

Recalling that $\ln 1 = 0$

$$\begin{bmatrix} x_1-3 & x_2-6 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x_1-3 \\ x_2-6 \end{bmatrix} = \begin{bmatrix} x_1-3 & x_2+2 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x_1-3 \\ x_2+2 \end{bmatrix} + \ln 4$$

$$\begin{bmatrix} x_1-3 & x_2-6 \end{bmatrix}\begin{bmatrix} 2(x_1-3) \\ 0.5(x_2-6) \end{bmatrix} = \begin{bmatrix} x_1-3 & x_2+2 \end{bmatrix}\begin{bmatrix} 0.5(x_1-3) \\ 0.5(x_2+2) \end{bmatrix} + \ln 4$$

$$2(x_1-3)^2 + 0.5(x_2-6)^2 = 0.5(x_1-3)^2 + 0.5(x_2+2)^2 + \ln 4$$

$$1.5x_1^2 - 9x_1 + 13.5 + 0.5[x_2^2 - 12x_2 + 36 - (x_2^2 + 4x_2 + 4)] - \ln 4 = 0$$

$$1.5x_1^2 - 9x_1 - 8x_2 + 29.5 - \ln 4 = 0$$

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?

Solution Exercise 4

a)

$$x = 4 + 0.7 \times 1 + 0.03 \times 0 + 0.4 \times (\frac{1}{3}) \times \log 40 = 0.7 + 4 + 0.5340 = 4.9136$$

$$p(x) = \frac{1}{1+e^{-4.9316}} = 0.9928$$

b)

$$\ln\frac{0.9928}{1-0.9928} = 4.9264$$

c) One can interpret that it shows the odds for somebody who has zero values for all X. But a more interesting interpretation would be it gives a baseline or background odds, which is the odds that result from logistic model without any X. Such an estimate can serve as a starting point for comparing other estimates of risk or odds when one or more X are considered.

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.

Solution Exercise 5

Step 1 Define number of neighbors. Suppose $K = 3$.

Step 2 Calculate distance between query pointy and training samples:

$(3, 7)$ from $(7, 7) : (7 - 3)^2 + (7 - 7)^2 = 16$
$(3, 7)$ from $(7, 4) : (7 - 3)^2 + (4 - 7)^2 = 25$
$(3, 7)$ from $(3, 4) : (3 - 3)^2 + (4 - 7)^2 = 9$
$(3, 7)$ from $(1, 4) : (1 - 3)^2 + (4 - 7)^2 = 13$

Step 3 Determine nearest neighbors based on k-th minimum distance.
The 3 nearest neighbors are $(3, 4), (1, 4), (7, 7)$. 2 Good , 1 Bad. Thus, query sample is classified as 'Good'.