COMP 473/6731, Fall 2019 **CONCORDIA UNIVERSITY**

PATTERN RECOGNITION
Assignment 1.
Bayesian Decision Theory.
Due: Sept. 29, 2019.

Do the following problems. [1]. It's on 3 hours reserve in the Webster Library

1. (25 marks) Problem 4, section 2.3, p. 66
   Consider the minimax criterion for a two-category classification problem.

   (a) Fill in the steps of the derivation of Eq. 23:

   $$R(P(\omega_1)) = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{\mathcal{R}_1} p(x|\omega_2)dx$$
   $$+P(\omega_1) \left[(\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{\mathcal{R}_2} p(x|\omega_1)dx - (\lambda_{12} - \lambda_{22}) \int_{\mathcal{R}_1} p(x|\omega_2)dx \right].$$

   (b) Explain why overall Bayes risk must be concave down as a function of the prior $P(\omega_1)$ as shown in Fig. 2.4

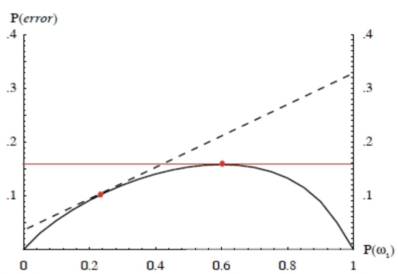

Figure 2.4: The curve at the bottom shows the minimum (Bayes) error as a function of prior probability $P(\omega_1)$ in a two-category classification problem of fixed distributions. For each value of the priors (e.g., $P(\omega_1) = 0.25$) there is a corresponding optimal decision boundary and associated Bayes error rate. For any (fixed) such boundary, if the priors are then changed, the probability of error will change as a linear function of $P(\omega_1)$ (shown by the dashed line). The maximum such error will occur at an extreme value of the prior, here at $P(\omega_1) = 1$. To minimize the maximum of such error, we should design our decision boundary for the maximum Bayes error (here $P(\omega_1) = 0.6$), and thus the error will not change as a function of prior, as shown by the solid red horizontal line.

   (c) Assume we have one-dimensional Gaussian distributions $p(x|\omega_i) \sim N(\mu_i, \sigma_i^2), i = 1, 2$, but completely unknown prior probabilities. Use the minimax criterion to find the optimal decision point x^* in terms of μ_i and σ_i under a zero-one risk.

   (d) For the decision point x^* you found in (c), what is the overall minimax risk? Express this risk in terms of an error function $erf(\cdot)$.

   (e) Assume $p(x|\omega_1) \sim N(0, 1)$ and $p(x|\omega_2) \sim N(1/2, 1/4)$, under a zero-one loss. Find x^* and the overall minimax loss.

   (f) Assume $p(x|\omega_1) \sim N(5, 1)$ and $p(x|\omega_2) \sim N(6, 1)$. Without performing any explicit calculations, determine x^* for the minimax criterion. Explain your reasoning.

2. (15 marks) Problem 8, section 2.3, p. 67
   Let the conditional densities for a two-category one-dimensional problem be given by the Cauchy distribution

   $$p(x|\omega_i) = \frac{1}{\pi b} \frac{1}{1 + \left(\frac{x - a_i}{b}\right)^2}, \quad i = 1, 2.$$

   (a) By explicit integration, check that the distributions are indeed normalized.

[1] see the textbook Duda, Hart, and Stork, *Pattern Classification*, Wiley, 2-nd edition, 2001

(b) Assuming $P(\omega_1) = P(\omega_2)$, show that $P(\omega_1|x) = P(\omega_2|x)$ if $x = (a1 + a2)/2$, i.e., the minimum error decision boundary is a point midway between the peaks of the two distributions, regardless of b.

(c) Plot $P(\omega_1|x)$ for the case $a_1 = 3, a_2 = 5$ and $b = 1$.

(d) How do $P(\omega_1|x)$ and $P(\omega_2|x)$ behave as $x \to -\infty$? $x \to \infty$? Explain.

3. (15 marks) Problem 27, section 2.6, p. 72

(a) Suppose we have two normal distributions with the same covariances but different means: $N(\boldsymbol{\mu_1}, \boldsymbol{\Sigma})$ and $N(\boldsymbol{\mu_2}, \boldsymbol{\Sigma})$. In terms of their prior probabilities $P(\omega_1)$ and $P(\omega_2)$, state the condition that the Bayes decision boundary *not* pass between the two means.

(b) Consider an example on pp. 45-46 of Lecture 3 with general normal distributions for a 3-class problem. For the given distributions derive the equations of the boundaries between classes (the plots of the boundaries are given in the notes). Compute numerically the coordinates of the two points where the boundaries of all three classes meet.

4. (20 marks) Problem 37 (c), section 2.8, p. 74
Consider a two-category classification problem in two dimensions with $P(\omega_1) = P(\omega_2) = 1/2$,

$$p(\mathbf{x}|\omega_1) \sim N\left(\mathbf{0}, \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right)$$

and

$$p(\mathbf{x}|\omega_2) \sim N\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\right).$$

(a) Calculate the Bayes decision boundary.

(b) Calculate the Bhattacharyya error bound.

5. (25 marks) Problem 2, p. 80 (computer exercise)
Use the classifier given by Eq. 49:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu_i})^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu_i}) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

to classify the following 10 samples from the table

	ω_1			ω_2			ω_3		
sample	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	−5.01	−8.12	−3.68	−0.91	−0.18	−0.05	5.35	2.26	8.13
2	−5.43	−3.48	−3.54	1.30	−2.06	−3.53	5.12	3.22	−2.66
3	1.08	−5.52	1.66	−7.75	−4.54	−0.95	−1.34	−5.31	−9.87
4	0.86	−3.78	−4.11	−5.47	0.50	3.92	4.48	3.42	5.19
5	−2.67	0.63	7.39	6.14	5.72	−4.85	7.11	2.39	9.21
6	4.94	3.29	2.08	3.60	1.26	4.36	7.17	4.33	−0.98
7	−2.51	2.09	−2.59	5.37	−4.63	−3.65	5.75	3.97	6.65
8	−2.25	−2.13	−6.94	7.18	1.46	−6.66	0.77	0.27	2.41
9	5.56	2.86	−2.26	−7.39	1.17	6.30	0.90	−0.43	−8.71
10	1.03	−3.33	4.33	−7.50	−6.32	−0.31	3.52	−0.36	6.43

in the following way. Assume that the underlying distributions are normal.

(a) Assume that the prior probabilities for the first two categories are equal $P(\omega_1) = P(\omega_2) = 1/2$ and $P(\omega_3) = 0$ and design a dichotomizer for those two categories using only the x_1 feature value.

(b) Determine the empirical training error on your samples, i.e., the percentage of points misclassified.

(c) Use the Bhattacharyya bound to bound the error you will get on novel patterns drawn from the distributions.

(d) Repeat all of the above, but now use *two* feature values, x_1, and x_2.

(e) Repeat, but use all three feature values.

(f) Discuss your results. In particular, is it ever possible for a finite set of data that the empirical error might be *larger* for more data dimensions?