Pattern Recognition - Fall 2024

Homework 1
Due date: Tuesday, November 05, 2024

1. **Section 2.1**
   (20 %) Consider a two-class classification problem. Given the class-conditional probability density functions

$$
\begin{aligned}
p(x|\omega_1) &\sim N(-1, 4) \\
p(x|\omega_2) &\sim N(2, 3)
\end{aligned}
$$

   with $P(\omega_1) = 0.6$ and $P(\omega_2) = 0.4$. Please specify the decision regions based on Bayesian decision theory.

2. **Section 2.4, Problem 14(b)**
   In many pattern classification problems one has the option either to assign the pattern to one of c classes, or to *reject* it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. Let

$$
\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \quad i, j = 1, \ldots, c \\ \lambda_r & i = c+1 \\ \lambda_s & \text{otherwise,} \end{cases}
$$

   where λ_r is the loss incurred for choosing the $(c+1)$th action, rejection, and λ_s is the loss incurred for making any substitution error.

   (b) (20 %) Plot the conditional risks and the decision regions for the two-category one-dimensional case having
   - $p(x|\omega_1) \sim N(1, 1)$
   - $p(x|\omega_2) \sim N(-1, 1)$
   - $P(\omega_1) = P(\omega_2) = 1/2$, and
   - $\lambda_r/\lambda_s = 1/4$.

3. **Section 2.6**
   Consider the problem of classifying 10 samples from Table 1. Assume that the underlying distributions are normal. For each category, the mean vector and the covariance matrix are given by

$$
\begin{aligned}
\boldsymbol{\mu} &= \frac{1}{10}\sum_{k=1}^{10} \mathbf{x}_k \\
\boldsymbol{\Sigma} &= \frac{1}{10}\sum_{k=1}^{10} (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^t
\end{aligned}
$$

   where $\mathbf{x}_k$ denotes the k-th samples in that category.

(a) (10 %) Assume that the prior probabilities for each category are as follows:

$$P(\omega_1) = 0.2, \quad P(\omega_2) = 0.3, \quad P(\omega_3) = 0.5$$

Determine mean vectors and covariance matrices for these three categories using x_1 and x_2 feature values.

(b) (10 %) Calculate the percentage of misclassified samples.

(c) (20 %) Repeat all of the above, but now use three feature values (i.e., $x_1, x_2,$ and x_3).

Table 1: Computer exercise 2 relies on the following data.

	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

4. **Section 2.8.3**
Suppose we have the data points and the corresponding labels stored in the file "roc_data.mat".

(a) (10 %) Write your own code to plot the receiver operating characteristic (ROC) curve.

(b) (10 %) Write your own code to determine the area under the ROC curve (often referred to as AUC).