

# Homework 4

CSE 802 - Pattern Recognition and Analysis

Instructor: Dr. Arun Ross

Points: 100 (+10 Bonus Points)

Due Date: April 22, 2020

---

Please read the following instructions carefully:

1. You are permitted to discuss the following questions with others in the class. However, you must write up your own solutions to these questions. Any indication to the contrary will be considered an act of academic dishonesty.
2. A soft-copy of this assignment must be uploaded in D2L by April 22, 12:40 pm. In this copy, please include the names of individuals you discussed this homework with and the list of external resources (e.g., websites, other books, articles, etc.) that you used to complete the assignment (if any). Late submissions will not be graded.
3. When solving equations or reducing expressions you must explicitly show every step in your computation and/or include the code that was used to perform the computation. Missing steps or code will lead to a deduction of points.
4. Code developed as part of this assignment must be included as an appendix to your submission or inline with your solution.

- 
1. Generate 100 random training points from *each* of the following two distributions:  $N(20,5)$  and  $N(35,5)$ . Write a program that employs the Parzen window technique with a **Gaussian kernel** to estimate the density,  $\hat{p}(x)$ , using *all* 200 points. Note that this density conforms to a *single bimodal* distribution.
    - (a) [15 points] Plot the estimated density function for each of the following **window widths**:  $h = 0.01, 0.1, 1, 10$ . [Note: You can estimate the density at **discrete values of  $x$  in the  $[0,55]$  interval with a step-size of 1.**]
    - (b) [10 points] Repeat the above after generating 500 training points from each of the two distributions, and then 1,000 training points from each of the two distributions.
    - (c) [5 points] Discuss how the estimated density changes as a function of the window width and the number of training points.
  2. Consider the dataset available [here](#). It consists of two-dimensional patterns,  $\mathbf{x} = [x_1, x_2]^t$ , pertaining to 3 classes ( $\omega_1, \omega_2, \omega_3$ ). The feature values are indicated in the first two columns while the class labels are specified in the last column. The priors of all 3 classes are the same and a 0-1 loss function is assumed. Partition this dataset into a training set (the first 250 patterns of each class) and a test set (the remaining 250 patterns of each class).

(a) [10 points] Let

$$\begin{aligned}p([x_1, x_2]^t | \omega_1) &\sim N([0, 0]^t, 4I), \\p([x_1, x_2]^t | \omega_2) &\sim N([10, 0]^t, 4I), \\p([x_1, x_2]^t | \omega_3) &\sim N([5, 5]^t, 5I),\end{aligned}$$

where  $I$  is the  $2 \times 2$  identity matrix. What is the error rate on the test set when the Bayesian decision rule is employed for classification? Report the confusion matrix as well.

(b) [15 points] Suppose  $p([x_1, x_2]^t | \omega_i) \sim N(\mu_i, \Sigma_i)$ ,  $i = 1, 2, 3$ , where the  $\mu_i$ 's and  $\Sigma_i$ 's are *unknown*. Use the training set to compute the **MLE of the  $\mu_i$ 's and the  $\Sigma_i$ 's**. What is the error rate on the test set when the Bayes decision rule using the *estimated parameters* is employed for classification? Report the confusion matrix as well.

(c) [15 points] Suppose the form of the distributions of  $p([x_1, x_2]^t | \omega_i)$ ,  $i = 1, 2, 3$  is unknown. Assume that the training dataset can be used to estimate the density at a point using the Parzen window technique (**a spherical Gaussian kernel** with  $h = 1$ ). What is the error rate on the test set when the Bayes decision rule is employed for classification? Report the confusion matrix as well.

(d) [10 points] Implement the 1-nearest neighbor (1-NN) method for classifying the patterns in the test set. What is the error rate of the 1-NN method on the test set? Report the confusion matrix as well.

3. [20 points] The **iris (flower) dataset** consists of 150 4-dimensional patterns belonging to three classes (setosa=1, versicolor=2, and virginica=3). There are 50 patterns per class. The 4 features correspond to (a) sepal length in cm, (b) sepal width in cm, (c) petal length in cm, and (d) petal width in cm. Note that the class labels are indicated at the end of every pattern.

Design a  $K$ -NN classifier for this dataset. Choose the first 25 patterns of each class for training the classifier (i.e., these are the prototypes) and the remaining 25 patterns of each class for testing the classifier. [Note: Any ties in the  $K$ -NN classification scheme should be broken at random.]

(a) In order to study the effect of  $K$  on the performance of the classifier, report the confusion matrix for  $K=1, 5, 9, 13, 17, 21$ .

(b) Plot the classification accuracy as a function of  $K$ . Discuss your observations.

4. [10 points] Based on the notation developed in class, write down the Sequential Backward Selection (SBS) algorithm and the Sequential Floating Backward Selection (SFBS) algorithm.

---