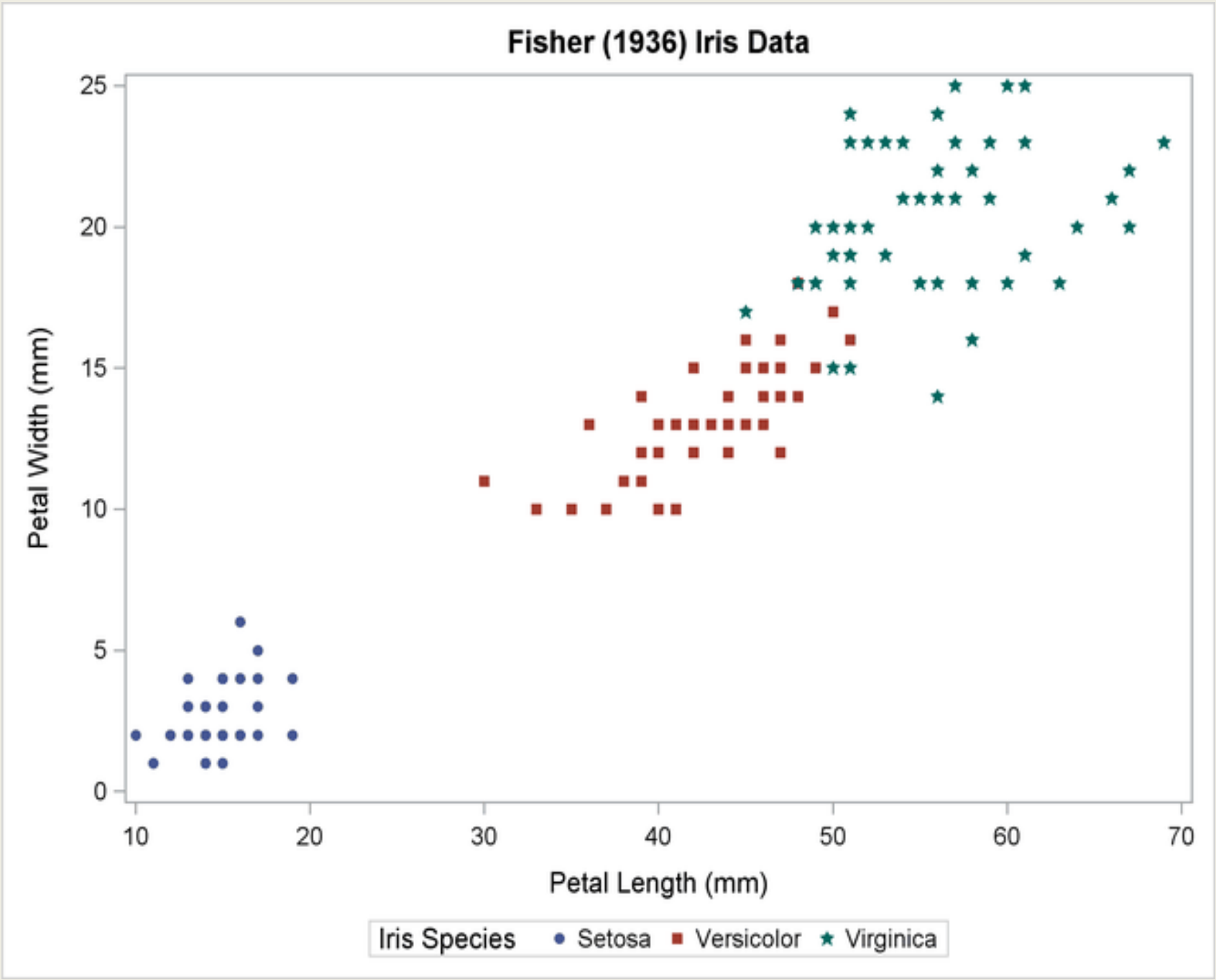
CISC 874 ASSIGNMENT 2 IMPLEMENT A HYBRID COMPETITIVE LEARNING NETWORK

Farhana Zulkernine

10 marks

General Instructions for Code and Submission (for all assignments)

1. You can use any programming language (preferred Python, Java – No GUI tool should be used). Python libraries can be used for matrix multiplication, producing charts or if needed, downloading data
2. Make one zip file named as Asgn2_studentID which should include
   a. *A PDF file (template given on the next page) containing the ANN design, accuracy in precision, recall, and confusion matrix*
   b. *Program code with comments in the code to explain what each of your program files and functions are for*
3. Upload zip file to the OnQ site
4. Mark will be deducted based on late policy (-1 per day after due date until the end date after which no assignments will be accepted)


Fisher (1936) Iris Data
Petal Width (mm)
25
20
15
10
5
0
10
20
30
40
50
60
70
Petal Length (mm)
Iris Species
Setosa
Versicolor
Virginica

Part 1

1. Design a Kohonen clustering network ANN model and train the model using simple competitive learning algorithm (do not use the label given in the data). Use the iris training data set. We call this KH1. **(3 marks)**
2. Train the model weights iteratively to find three clusters in the dataset such that the clustering error is minimum. Use appropriate terminating conditions**. Note down the cluster centroid numbers against the class labels, cluster centroid values (weight vectors), and the total clustering error**. Because the labels are given, the class labels should ideally correspond to the cluster centroids although this may not happen. **Explain your results in the report**.
3. Test how the network performs on the test.txt data to see if the data points are placed in the correct cluster i.e., grouped in the same cluster with the other data points in the same class. **Note down cluster number for each test data and total error to add to the report file**.

Kohonen Network from the lecture slide

$$\text{Output} = \begin{cases} 1 & \text{if node is a winner} \\ 0 & \text{otherwise} \end{cases}$$

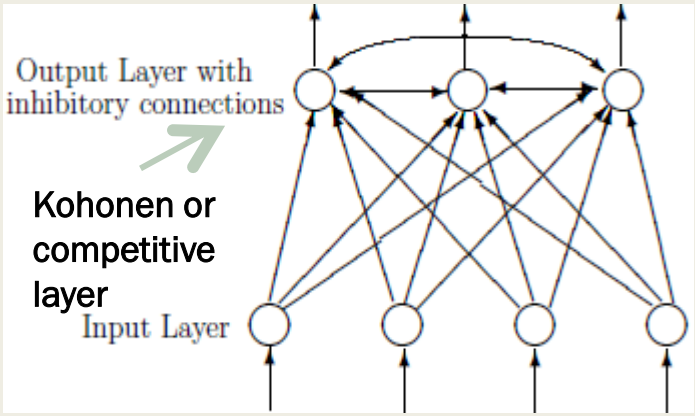


Figure: A simple competitive learning network

Part 2

4. Implement a PCA network to reduce the dimension of the iris training dataset from 4 to 3 (or 2). **Note down the network weights to add to the report file. Write the output having reduced dimension to a train1.txt and test1.txt file.**
5. Now feed the output from the PCA to a Kohonen network version KH2 (has less than 4 inputs). Train this integrated PCA-KH2 network with train.txt to **find 3 cluster centroids and total error for the 3 clusters.** Here you will be training weights only for the KH2 part. *Hint:* You can use the stored train1.txt to train only KH2 instead of feeding the data live from PCA.
6. Validate PCA-KH2 using test1.txt to see if the data points are placed in the right cluster. **Note the clusters for each test1.txt data point for reporting and the total clustering error.**

Mark Distribution – Code 7 Marks and PDF report 3 marks

- For fully functional executable codes with comments you will get full marks. For partially correct code you will get less marks.
  - *KH1 - 3 marks*
  - *PCA – 2 marks*
  - *PCA-KH2 – 2 marks*
- For parts 1 and 2 of the assignment list the following ***for both training and test datasets*** **for KH1 and PCA-KH2**
  - *Report the consistency between class label and cluster number (points in the same class should be placed in the same cluster) using a confusion matrix for train+test and train1+test1 data. (1 mark)*
  - *Report the centroids for KH1 and PCA-KH2, the weight matrix of PCA and the total clustering error for (E1) train+test, and (E2) train1+test1 data points (1 mark)*
  - *The observation and reasoning for the inconsistency of class label and clusters, and clustering error with data of reduced dimension of KH1 and PCA-KH2. (1 mark)*