Habib University
CSE 351 - Artificial Intelligence
Fall 2021
Assignment # 2 – Machine Classification (Classification)

Objective:

This assignment gives students hands-on experience of learning and evaluating classifiers and familiarize them with some commonly used machine learning libraries in python.

Q1 – Text Classification using Naïve Bayes [25 points]

You are given a dataset of textual summary of medical queries classified into five different categories. You have to build your own naïve Bayes classifier to predict these categories for future queries. The task includes:

- **[15 points]** Implement Learn(..) method that takes training and test data (in the form of word vectors) and learns a Naïve Bayes classifier (i.e. all probabilities that a Naïve Bayes classifier needs). You have to do your own implementation of Naïve Bayes here.
- **[05 points]** Some conditional probabilities may turn out to be zero in the training dataset. Modify your classifier such that it applies Laplacian smoothing while learning conditional probabilities to avoid such zero values.
- **[05 points]** Implement Predict(…) that takes a learned classifier and test data and returns predicted values.
- **[10 points]** Implement Evaluate(…) that takes actual and predicted labels and returns precision, recall, f-measure. The method will also display confusion matrix.

The code will be written in the attached Ass2_classification.ipynb. In addition to the lecture slides of Naïve Bayes classifier, you can also refer to this link for further details on using Naïve Bayes for text classification.

Q2 – Image Classification in Fashion Retail [15 Points]

You are given a Fashion-MNIST dataset comprising of 28x28 grayscale images of 60,000 fashion products from 10 categories, with 7,000 images per category. The training set has 60,000 images and the test set has 10,000 images. Some sample images are shown below:

The dataset is available here.

You have to build a scikit-learn based classifier for this dataset to classify each input image into one (out of 10) categories. You are required to experiment with three to four different classifiers and report accuracy of each of them.

Note: Your code will follow the same structure given in Problem 1.

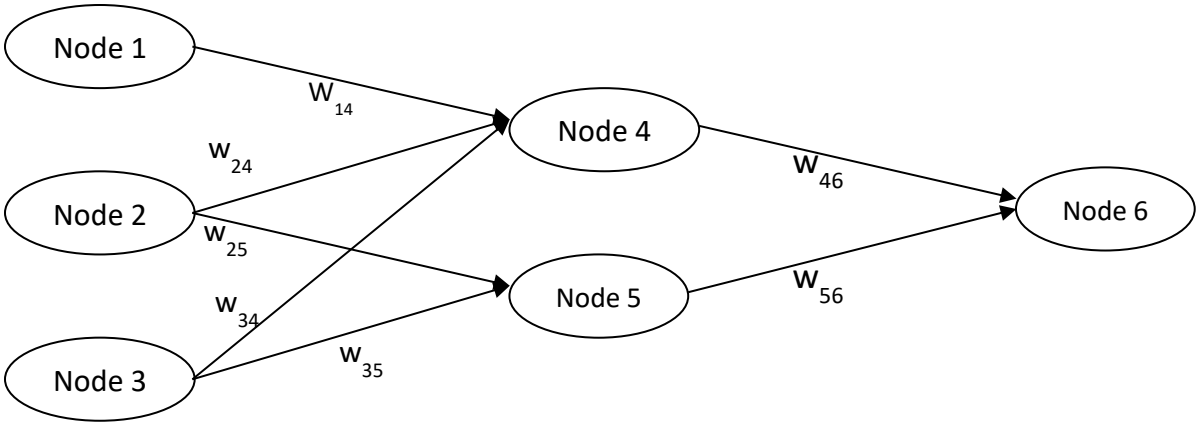
Q 3 – Forward and Back propagation through ANN [10 points]

You have to manually run one complete iteration of feed-forward and back-propagation over the network given below. At the end of this iteration, you will get updated weights to be used in the next iteration. Sigmoid function will be used as activation function.

w_{14}=0.35, w_{24}=0.15, w_{25}=-0.10, w_{34}=-0.20, w_{35}=0.20, w_{46}=0.40, w_{56}=0.25

Actual Output(T)= 0.8, Learning Rate (r) = 0.8

Input : (0.5, 0.3,0.9)



You are required to show your working. The result will be given in the following tabular format:

Input of Node 4		Error at Node 6	
Output of Node 4		Error at Node 4	
Input of Node 5		Error at Node 5	
Output of Node 5		Updated w_{46}	
Input of Node 6		Updated w_{56}	
Output of Node 6		Updated w_{14}	
		Updated w_{24}	
		Updated w_{34}	
		Updated w_{25}	
		Updated w_{35}	

Submission Instructions

The assignment will be submitted on the LMS by the due date (announced on LMS). No email submission will be accepted. The submitted file should be in the form of a ZIP file named as **<studentid>_Ass2** containing separate folders for each question (source code for Q1 and Q2 and a pdf file for Q3). The source code of Q1 and Q2 must follow the structure given in the attached notebook.