HOMEWORK 2

The homework has four parts. Each part must be done separately.

Write a program that classifies the loan applicants of a bank. The program uses the nearest neighbor algorithm. The applicants are classified based on the credit score, income, age, sex, and marital status. The applicants are classified into four classes namely high risk, medium risk, low risk, and undetermined. The training data is obtained from the past loans.

[671 89 55 male married | high]
[750 60 42 female single | low]
[810 70 60 male divorced | ?]

The program uses an appropriate number of nearest neighbors and a distance measure. The program performs appropriate data conversion and normalization. The program uses the leave one out method for validation. The program reads the training data and the test data from two user specified files. The program writes the classified data to a user specified file. The program displays the validation error on the screen.

Run the program on the given training file and the test file. Submit the classified data, the validation error, and the number of nearest neighbors. All three information must be in one file.

Write a program that recognizes the images of digits. Specifically the program tells whether an image is one or zero. An image is a 16×16 array of bits. The program uses the nearest neighbor algorithm. The images are classified based on their 256 bits. The images are classified into two classes namely one and zero.

```
0 0 0 0 0 0 0 0 0 0        0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 1 0 0 0 0        0 1 1 1 1 1 1 1 0 0
0 0 0 0 1 1 0 0 0 0        0 1 1 0 0 0 0 0 1 0
0 0 0 1 1 1 1 0 0 0        0 0 1 0 0 0 0 1 1 0
0 0 0 0 1 1 0 0 0 0        1 1 0 0 0 0 0 1 1 0
0 0 0 0 1 0 0 0 0 0        0 1 1 0 0 0 0 0 1 0
0 0 0 0 1 1 0 0 0 0        1 1 0 0 0 0 0 1 1 0
0 0 0 1 1 0 0 0 0 0        0 1 0 0 0 0 0 1 0 0
0 0 0 0 1 1 0 0 0 0        0 0 1 1 1 0 1 1 0 0
0 0 0 0 0 0 0 0 0 0        0 0 0 0 0 1 0 0 0 0
```

The program uses an appropriate number of nearest neighbors and a distance measure. The program uses the leave one out method for validation. The program reads the training data and the test data from two user specified files. The program writes the classified data to a user specified file. The program displays the validation error on the screen.

Create a training file containing at least twenty images, some ones and some zeros. Each training record contains one image and its label. Similary create a test file containing at least five images, some ones and some zeros. Each test record contains one image. Run the program on the training file and the test file. Submit the classified data, the validation error, and the number of nearest neighbors. All three information must be in one file. Also submit the training file and the test file.

Write a program that decides whether the applicants for a Java programming job are interviewed or not, by looking at their resumes. The program uses the Bayes algorithm. The program classifies the resumes based on the knowledge of languages, Java knowledge, years of experience, computer science degree, and academic performance. The resumes are classified into two classes namely interview and no interview. The training data is obtained from the past hiring experience.

[1 java 2 other B | interview]
[0 no 1 cs C | no]
[2 no 2 cs B | ?]

The program performs appropriate data conversion. The program uses the leave one out method for validation. The program reads the training data and the test data from two user specified files. The program writes the classified data to a user specified file. The program displays the validation error on the screen.

Run the program on the given training file and the test file. Submit the classified data and the validation error. Both information must be in one file.

Write a program that groups the customers of a bank into clusters. These clusters help the bank to identify various types of customers, and thereby improve customer service, communication, and marketing. The program uses the k-means algorithm. The customers are clustered based on the age, income, and credit score. The number of clusters and the cluster labels are unknown.

[27 65 521]
[38 45 750]
[70 30 632]

The program uses an appropriate number of clusters and a distance measure. The program performs appropriate data conversion and normalization. The program reads the records from a user specified file. The program writes the cluster information to a user specified file. This file contains the records and their cluster labels in a grouped format.

Run the program on the given record file. Submit the records and their cluster labels in a grouped format. Also submit the number of clusters. Both information must be in one file.

Program must be written Java language. Write comments explaining all steps. Run the program on the data provided separately. Submit the program and the results.