AINT 252 Computational Theory and AI

Coursework 2018/2019

The deadline for coursework submission: Wednesday 8th May, 2019, 15:00

Please submit electronically via DLE system

There are three questions in this coursework. For the coursework submission you have to select any **TWO** questions and submit answers to these **two selected questions only**.

Your submission: one PDF file which contains:

1) **Reports (answers) for two selected questions**
2) **All MATLAB codes which you have developed to answer two selected questions**

This coursework is an individual work (NOT a group work)

Therefore, the coursework should be **your own work.** Make sure that the **report** is **your own work** and you understand what has been done.

In this coursework you will practice how to use different methods for data classification. There are three questions in this coursework which include different AI algorithms. During your Labs you will practice how to use different methods.

QUESTION 1. K-means clustering (unsupervised learning).

DATA for QUESTION 1. To get data for k-means clustering, use the MATLAB code "gen_kmeansdata.p". This code generates your personal data matrix. Copy this code to your folder, which you use for your MATLAB scripts. Use the MATLAB command: X=gen_kmeansdata(ID), where ID is your student number. The data matrix X contains four columns representing features and N rows representing objects (cases).

Data analysis.

1) **Report N** - the total number of rows (objects, cases) in your data.
2) For each column (feature) from 1 to 4 **report**: the mean, the standard deviation and the histogram.
3) **Report** the covariance matrix (4x4) and the correlation matrix (4x4).
4) **Report** a short (less than 0.5 page) description of your data analysis.

Marks 10

K-means algorithm: find the optimal number of classes

Use k–means MATLAB function "kmeans" to classify your data between k classes. To characterise the algorithm performance use the MATLAB function "silhouette".

a) Repeat the following procedure three times: Select the value of parameter k (k=3, 4, 5) and execute the MATLAB function for clustering ("kmeans") and the MATLAB function for result representation ("silhouette");

   Report: 1) Coordinates of centres of k classes
   2) The silhouette plot
   3) The mean silhouette measure

Marks 30

b) Compare the results of three classifications (k=3, 4, 5) and select the best classification using the mean silhouette measure.

   1) **Report** the optimal number of classes (Ko).
   2) **Visualization**: for the optimal number of classes Ko plot the 3D projection of the clustered data, use different colours to show objects of different classes, also in the same figure plot projections of centres of clusters; **report** the MATLAB figure.

Marks 10

TOTAL MARKS FOR QUESTION 1: 50 MARKS

QUESTION 2. K Nearest Neighbour (KNN) classifier for supervised learning

DATA for QUESTION 2.

To get data for your unsupervised classifiers, use the MATLAB code "gen_superdata.p". Copy this code to your folder which you use for your MATLAB scripts. To get your personal data matrix use the MATLAB command: Y=gen_superdata(ID), where ID is your student number. The data matrix Y contains N rows representing objects (cases) and six columns: five columns (1-5) represent features and the sixth column contains class labels (1,2,...).

Data analysis.

1) **Report N -** the total number of rows (objects, cases) in your data.
2) For each column (feature) from 1 to 5 **report**: the mean and the standard deviation.
3) **Report** the covariance matrix (5x5) and the correlation matrix (5x5).
4) **Report** the number of classes.
5) **Report** a short (less than 0.5 page) description of your data analysis.

Marks 10

Data Pre-processing: Prepare the training set: **randomly** select 60% of rows (object, cases) from the matrix Y. All other rows will be used for testing.

Marks 5

KNN classifier.

a) Training. Use the training data set to train the KNN classifier. Testing. Select K=5 and test the classifier performance using the testing data. **Report** the result of testing using the confusion matrix for testing data set with K=5.

Marks 15

b) Investigate how the classification result depends on K: Repeat the same procedure of testing as in Q 2 a) using K=7. For the testing data set with K=7 **report** the result of testing using the confusion matrix.

Marks 15

c) Using the overall percentage of correct classifications of testing data, compare two KNN classifications: for K=5 and for K=7 and **report** your conclusion.

Marks 5

TOTAL MARKS FOR QUESTION 2: 50 MARKS

QUESTION 3. **Feedforward Neural Network (Multi-Layer Perceptron MLP) and Support Vector Machine (SVM) for supervised learning.**

DATA for QUESTIONS 3.

To get data for your unsupervised classifiers, use MATLAB code "gen_superdata.p". Copy this code to your folder which you use for your MATLAB scripts. To get your personal data matrix use the MATLAB command: Y=gen_superdata(ID), where ID is your student number. The data matrix Y contains N rows representing objects (cases) and six columns: five columns (1-5) represent features and the sixth column contains class labels (1,2,...).

Data analysis.

1) **Report N** - the total number of rows (objects, cases) in your data.
2) For each column (feature) from 1 to 5 **report**: the mean and the standard deviation.
3) **Report** the covariance matrix (5x5) and the correlation matrix (5x5).
4) **Report** the number of classes.
5) **Report** a short (less than 0.5 page) description of your data analysis.

Marks 10

Data Pre-processing: Prepare the training set: **randomly** select 60% of rows (object, cases) from the matrix Y. All other rows will be used for testing.

Marks 5

MLP:

a) Training. Use the training data set to train the MLP classifier. **Report** important characteristics of MLP classifier (the number of hidden layers, the number of hidden neurons, etc). **Report** the percentage of correct classifications for the training data.
b) Testing. Using the testing data, test the performance of MLP classifier. **Report** the percentage of correct classifications for the testing data.

Marks 15

SVM:

c) Training. Use the training data set to train the SVM classifier. **Report** important characteristics of the SVM classifier (kernel function, loss function, etc). **Report** the percentage of correct classifications on the training data.
d) Testing. Using the testing data, test the performance of SVM classifier. **Report** the percentage of correct classifications for the testing data.

Marks 15

Comparison MLP and SVM:

Compare results of training and testing of two classifiers. **Report** your conclusions about the performance of MLP and SVM classifiers.

Marks 5

TOTAL MARKS FOR QUESTION 3: 50 MARKS