**DS Lab**

**Lab 1**

**Question 1**

Write a python program to sort a [list](http://14.139.187.105/moodle/mod/resource/view.php?id=3108) of ‘k’ elements using insertion sort.

Ensure the proper boundary conditions.

Note: sort() [function](http://14.139.187.105/moodle/mod/resource/view.php?id=3109) should not be used.

**Question 2**

Write a python program to search for a value ‘y’ in a [list](http://14.139.187.105/moodle/mod/resource/view.php?id=3108) of ‘k’ elements using binary search.

Ensure proper boundary conditions

Note: sort() [function](http://14.139.187.105/moodle/mod/resource/view.php?id=3109) can be used to sort the [list](http://14.139.187.105/moodle/mod/resource/view.php?id=3108)

**Question 3**

Write a python program to search for a value ‘y’ in a [list](http://14.139.187.105/moodle/mod/resource/view.php?id=3108) of ‘k’ elements using binary search.

Ensure proper boundary conditions

Note: sort() [function](http://14.139.187.105/moodle/mod/resource/view.php?id=3109) should not be used to sort the [list](http://14.139.187.105/moodle/mod/resource/view.php?id=3108)

**Lab 7**

**Question 1**

Read the given "[diabetes.csv](http://14.139.187.105/moodle/mod/resource/view.php?id=3454)" into a pandas data frame. Write a menu-driven program to calculate and print the following,

1. Find the probability of diabetes given the dataset. Also, calculate the probability of diabetes given

a) Age above 50

b) Age between 40 and 50

c) Age between 30 and 40

d) Age less than 30

**Question 2**

Find the probability of diabetes with a glucose level of more than 120 + blood pressure of more than 90 + skin thickness of more than 30 + insulin above 150 + BMI above 25.

Note: In the outcome column given in the dataset, 1 means that diabetes is present, and 0 means the absence of diabetes

**Lab 7**

**Question 1**

Classify whether a person is diabetic or not using (a) Minimum distance classifier (Supervised Learning) (b) K- means algorithm (Unsupervised Learning) from the given diabetic.csv file.

Implement the above using a python program without using built-in functions to implement (a) and (b).

Note:

* Outcome is the class label (1 - is diabetic and 0 - is normal. Only for supervised learning)
* Divide the data into train and test (80% and 20%)
* Cluster the attributes using K-means algorithm and predict the outcomes on the test data
* Classify the attributes using Minimum distance classifier and predict the outcomes on the test data

**Question 2**

The given 'diabetes1.csv' file has diabetic information about female patients. The features/attributes in the file are as follows,

Pregnancies: Number of times pregnant

Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test

BloodPressure: Diastolic blood pressure (mm Hg)

SkinThickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mu U/mL)

BMI: Body mass index (weight in kg/(height in m)^2)

DiabetesPedigreeFunction: Diabetes pedigree [function](http://14.139.187.105/moodle/mod/resource/view.php?id=3109)

Age: Age (years)

Outcome: Class variable (0 - non-diabetic or 1 - diabetic)

Write a python program (with pandas) to read the given data and display the following:  
1. Mean, median, mode, minimum, maximum, and standard deviation for all the attributes excluding the attribute ‘Outcome’. Find the value of the correlation coefficient for ‘Age’ with all other attributes (excluding ‘Outcome’), and ‘BMI’ with all other attributes (excluding ‘Outcome’). Then, obtain the scatter plot between   
            a. ‘Age’ and each of the other attributes, excluding ‘Outcome’   
            b. ‘BMI’ and each of the other attributes, excluding ‘Outcome’ (You can use matplotlib library).