**Homework 2:**

**Answer:**

import pandas as pd

from sklearn import preprocessing

import numpy as np

pd.set\_option('display.precision',4)

from scipy.spatial import distance

from scipy.spatial.distance import pdist, squareform

#data extraction and describe the properties of the data

df\_red = pd.read\_csv('E:\ECE 657A\winter 2019\homework\hw2\winequality-red.csv',sep=";")

df\_white = pd.read\_csv('E:\ECE 657A\winter 2019\homework\hw2\winequality-white.csv', sep=";")

des\_red= df\_red.describe()

des\_white= df\_white.describe()

#saving the properties in excel

writer = pd.ExcelWriter('output.xlsx')

des\_red.to\_excel(writer,'Sheet1')

des\_white.to\_excel(writer,'Sheet2')

writer.save()

#histogram the data

#df\_red.hist()

#df\_white.hist()

#extracting first 10 rows

df\_red2=df\_red[0:10]

df1 = df\_red2.loc[:,'fixed acidity':'alcohol']

print(df1)

x = df1.values #returns a numpy array

**#anwer to ques no: 1**

#z-core scaling of data

std\_scale = preprocessing.StandardScaler().fit(x)

df\_std = std\_scale.transform(x)

df\_std =pd.DataFrame(df\_std)

des1=df\_std.describe()

print(des1)

des1.to\_excel(writer,'Sheet3')

writer.save()

#analysis only showing for wine red

#min-max scaling of data

minmax\_scale = preprocessing.MinMaxScaler().fit(x)

df\_minmax = minmax\_scale.transform(x)

df\_minmax = pd.DataFrame(df\_minmax)

des2=df\_minmax.describe()

print(des2)

des2.to\_excel(writer,'Sheet4')

writer.save()

#mean-subtracted scaling of data

meansub\_scale=df1-df1.mean()

des3=meansub\_scale.describe()

print(des3)

des3.to\_excel(writer,'Sheet5')

writer.save()

#similar analysis can be carried out for wine white

**#answer to question no:2**

#red wine data

#converting data into matrices

x1=df\_std.values

x2=df\_minmax.values

x3= meansub\_scale.values

#calculating distance matrix

y1= pdist(x1,'euclidean')

y2=squareform(y1)

#minimum of each array from distance matrix

t1= y2.min(axis=1)

#maximum of each array from distance matrix

t2= y2.max(axis=1)

#similar can be done for manhattan distance and cosine distance

#repeat the same for white wine data