

**B.Sc.(H) Computer Science**  
**Semester VI**  
**Practical Exam Jan-April 2020**  
**Machine Learning**

**All questions are compulsory.**

**The time duration is 3 hours.**

**Your answers must include .py files named ques1.py and ques2.py for Q1 and Q2 respectively. You also have to upload a word document saved with your name which should contain your name, College RollNo and also outputs of the two questions.**

**Q1)** Boston house prices dataset (already available in Scikit Learn) has the following characteristics:

Number of Instances: 506

Number of Attributes: 13 numeric/categorical predictive. Median Value (attribute 14) is usually the target.

Attribute Information (in order):

CRIM per capita crime rate by town, ZN proportion of residential land zoned for lots over 25,000 sq.ft., INDUS proportion of non-retail business acres per town, CHAS Charles River dummy variable (= 1 if tract bounds river; 0 otherwise), NOX nitric oxides concentration (parts per 10 million), RM average number of rooms per dwelling, AGE proportion of owner-occupied units built prior to 1940, DIS weighted distances to five Boston employment centres, RAD index of accessibility to radial highways, TAX full-value property-tax rate per \$10,000, PTRATIO pupil-teacher ratio by town, B  $1000(B_k - 0.63)^2$  where  $B_k$  is the proportion of blacks by town, LSTAT % lower status of the population, MEDV Median value of owner-occupied homes in \$1000's.

Missing Attribute Values: None

Build a linear regression model using Scikit Learn that predicts the house prices based on multiple features (all 13). You must report the performance of the model using the metrics: R-square and RMSE. Also, print the regression coefficients obtained.

**Q2)** Iris plants dataset (already available in Scikit Learn) has the following characteristics:

Number of Instances: 150 (50 in each of three classes)

Number of Attributes: 4 numeric, predictive attributes and the class

Attribute Information: sepal length in cm, sepal width in cm, petal length in cm, petal width in cm

Class: Iris-Setosa, Iris-Versicolour, Iris-Virginica

Missing Attribute Values: None

Write a program using Scikit Learn that utilizes Logistic regression to build a classification model using all the four features to predict the class of a plant. Print the confusion matrix, accuracy, precision and recall for the model.

Also, build a classification model in Scikit Learn using Neural Networks using all the features to predict the class a plant belongs to. Print the confusion matrix, accuracy, precision and recall for the model and compare its performance with the model created using Logistic regression.