Machine Learning

Assignments

Instructor: Dr. Swarup Chattopadhyay

Submission Due Date: 18-04-2023

Use the Python language to implement the following assignments and submit them by the due date.

Assignment-1:

Consider the Iris ML data set (Download it from University of California–Irvine Machine Learning Data Repository). Generate the variance covariance matrix (Matrix Dimension should be 4 cross 4, since the data has only 4 attributes). Then find out the first two principal components of it.

Assignment-2:

Read the following data sets in Python and fit the linear regression (simple and multiple) model over it. Estimates the parameters [a and b for one predictor and one dependent variables (Y=aX+b); a, b, and c for two predictors and one dependent variables (Y=aX1+bX2+c)]. Calculate the performance measures viz. SSE, SST, R^2, AdjR^2, MSE, RMSE values. Finally show iterative regression lines gradient descent optimization method for the following data sets. (Following data sets are publicly available in internet and/or UCI-ML, Kaggle repository)

i. Salary_data.csv
ii. Credit_Card_Expences.csv
iii. Real estate.csv

Assignment-3:

Download and/or upload Breast cancer data set in python (`from sklearn.datasets import load_breast_cancer`) and Fit logistic regression model over it. Calculate the measures of performances (Precision, Recall, Accuracy).

Assignmnt-4:

(i) Download and/or upload Iris classification data set in python (`from sklearn import datasets (), then iris = datasets.load_iris()`). It has Four features (Sepal Length, Sepal Width, Petal Length and Petal Width) and three class labels (Setosa, Versicolour, and Virginica).

(ii) Generate Synthetic Classification data set having 2000 sample points, three (3) different classes and 15 features using make_classification() in Python.

Then apply Minimum Distance Classifier, K-NN classifier, Decision Tree Classifier, SVM Classifiers, Bagging Classifier, Boosting Classifier models and calculate the measures of performances (viz. Precision, Recall, F-Measure, CV-Score), and plot AUC-ROC curve, and respective confusion matrices.

Assignmnt-5:

(i) Generate Synthetic clustering datasets dataset of groups in 2-dimensional space in python using make_blobs() function.

(ii) Consider the Iris data set also for clustering.

Then apply K-means, Hierarchichal clustering algorithms to find out optimal clusters present in the data sets. Calculate the measure of performances (NMI, F-Measure). Plot the Elbow method to determine number of clusters.

Assignmnt-6:

Implement Single layer Perceptron algorithm (SLP) to find out the linear separation boundaries of the following data sets. Update weight and bias parameters iteratively and draw the separation lines at each step.