National Institute of Electronics and Information Technology, Calicut

Certified AI Professional

Machine Learning

Assignment 8

(Total Marks 40)

(Submit py or ipynb file. All questions carry 4 marks each)
All are ML Model development. Required data exploration (like printing few rows of data, column names, printing target values etc.) is expected before proceeding to ML model development.

Clustering

1. Prepare an ML model using KMeans algorithm to cluster some sample input generated using make_blob function. Plot the clusters in different colours.
2. Cluster IRIS dataset on the basis of Sepal Length and Sepal Width and visualise them.
3. Do clustering with some suitable datasets in the link. http://cs.joensuu.fi/sipu/datasets/ (Include the program and link to the dataset also in the submission)

Linear Regression

4. Develop a Linear Regressor for Advertising.csv dataset (https://github.com/marcopeix/ISL-linear-regression/blob/master/data/Advertising.csv) and print the importatnt metrics for performance evaluation (MAE, MSE, RMSE & R2SCORE)

5. Implement an ML model for the bikeshare.csv dataset Regression Problem using Linear Regression (https://www.kaggle.com/c/bike-sharing-demand/data) Evaluate the model by splitting the data using train_test_split function. Compute Mean Squared Error, also plot the actual values Vs predictions graph

Logistic Regression

6. For the Breast Cancer dataset in sklearn
   a) Develop an ML Model using Logistic Regression
   b) Apply scaling on input columns before model development
   c) Apply normalization on input columns before model development

   (Print RMSE for all the three models)

Label Encoding

7. Implement a KNN Classifier for the banking.csv dataset. Use only the numeric columns from the dataset as input. Print accuracy score

8. Implement a KNN Classifier for the banking.csv dataset. Convert the categorical columns to numeric using Label Encoding and use those columns also in the input. Print accuracy score

SVM

9. Develop an SVM Classifier (SVC) for the Wine recognition dataset in sklearn. Print Classification Report

10. Develop an SVM repressor (SVR) to predict disease progression in Diabetes dataset