School of Computer Science Engineering and Technology

Course- BTech   Type- Core
Course Code- CSET301   Course Name-AIML
Year- 2022   Semester- Odd
Date- 12-09-2022   Batch- V Sem

Lab Assignment No. 4 . 1 _ 2

Exp. No.	Name	CO-1	CO-2	CO-3
4.1_2	Logistic Regression	✓	✓	--

Objective: To implement Logistic Regression Model (Classification Model)

Download the dataset from https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset?select=Bank+Customer+Churn+Prediction.csv (10)

This dataset is for ABC Multistate bank with following columns:

1. customer_id, unused variable.
2. credit_score, used as input.
3. country, used as input.
4. gender, used as input.
5. age, used as input.
6. tenure, used as input.
7. balance, used as input.
8. products_number, used as input.
9. credit_card, used as input.
10. active_member, used as input.
11. estimated_salary, used as input.
12. churn, used as the target. 1 if the client has left the bank during some period or 0 if he/she has not.
.

1. **Load** the data and print first 10 and last 10 rows using a suitable function. (5)

2. **Data Pre-processing step: (40)**
   a) Check the presence of missing values. Handle it if present.
   b) Check the presence of Categorical columns. Handle it if present. i.e., Transform categorical features into numerical features. (Hint: Use eitherone hot encoding, label encoding or any other suitable pre-processing technique).
   c) Scale the numerical columns value using minmax_scale() or any other scaling function.

3. Define **X** matrix (independent features) and **y** vector (target feature). (5)

4. **Split** the dataset into **80% for training** and rest **20% for testing** (sklearn.model_selection.train_test_split function) (5)

5. **Train** Logistic Regression Model using built-in function on the training set (sklearn.linear_model.LogisticRegression class). (10)

6. Use the trained model to **predict** on the **test set** and then (15)
    a. Print 'Accuracy' obtained on the testing dataset i.e. (sklearn.metrics.accuracy_score function)
    b. Confusion matrix (sklearn.metrics.confusion_matrix),
    c. Precision, Recall and F1 scores (sklearn.metrics.precision_recall_fscore_support)

7. Compare and analyse the **test accuracy** for different train-test splits of data such as 60-40, 70-30,80-20 and 90-10 with the help of **suitable graphs**. (15)

Suggested Platform: Python: Jupyter Notebook/Azure Notebook/Google Colab.