Machine Learning: Lab 3 – Credit Card Fraud Detection and ANNs

Download the Credit Card Fraud Detection dataset from https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Prerequisites: Python basics, numpy, pandas, matplotlib, seaborn, sklearn, etc.

Importing Data:

1. Randomly shuffle the dataset by taking a random seed of "42". Create a testing set from the last 20% rows of the dataframe (these must be the same for all the students). The remaining rows will be the training + validation set, with training : validation ratio of 80% : 20%. Determine

A) number of rows in training, validation and test sets, along with the structure, datatypes and value counts of the dataframes.

Data Cleaning:

1. Analyse the data.

2. Check for missing values and logically impute the dataset. Normalize the columns.

Classification:

1. Train a logistic regression model on the training set partition by taking all the features. Calculate the error on the validation set.

2. Train a neural network model (MLPClassifier) by taking all the features and predicting the result. Fix the random_state for training. Choose the solver as 'adam', and set the number of hidden layers to (10, 2). Vary the hidden layers to find the best set of results on validation set. Explore different training parameters of MLP.

3. Find the evaluation scores by constructing the confusion matrix (True Positive, False positive, True Negative and False Negatives). Precision = TP/(TP+FP), Recall = TP/(TP+FN)

4. Plot the precision and recall values on the same plot for different training iterations for the MLP model by varying max_iter as 25, 50, 75, ….

5. Make predictions on the test set by taking 3 of your best models. Report these 3 accuracy values.