Practical - 03.

* Aim: Given a bank customer, build a neural network based classifier that can determine whether they will leave or not in the next 6 months. Dataset: Description The case study is from an open source dataset from Kaggle. The dataset contains 10,000 sample points with 14 distinct features such as Cust id, credit score, geography, gender, age, tenure, Balance, etc.
Link to kaggle project: perform steps –
1. Read the dataset.
2. Distinguish the feature and target set and divide the data set into training and test sets.
3. Normalize the train and test data.
4. Initialize and build the model. Identify the points of improvement and implement the same.
5. Print the accuracy score and confusion matrix

THEORY:

1. Data set understanding: –

- The dataset is an open source bank customer churn dataset from Kaggle.
- It contains 10,000 samples with 14 features

2. Feature and target selection –
- Features (x) = All columns except CustomerID and exited.
- Target y = Exited column, indicating customer churn

3. Data pre processing -

- Train Test split -
  - Divide the dataset into training and testing sets 80% to 20% ratio.
  - Training set is used for model training, test set for evaluation.
- Normalization -
  - Scale numeric features using Standard Scaler or Min Max Scaler to improve neural network convergence.
  - Avoid scaling categorical colums or encode them first.

4. Neural network model -

- Architecture -
  - Input layer : Number of neurons equal to number of features.
  - Hidden layers : 1-3 layers with ReLU activation (experiment to improve performance).
  - output layer - Single neuron with sigmoid activation for binary classification.
- Compilation -
  - Loss function - binary - Crossentropy
  - Optimizer - adam (adaptive learning rate).
  - Metrics - accuracy.

- Training -
  - Use batch size of 32 or 64
  - Number of epochs : 50 - 100 (depending on convergence).
  - Include early stopping to prevent overfitting.
- Improvement techniques -
  - Add more hidden layers or neurons.
  - Adjust learning rate.
  - Apply dropout layers for regularization.
  - Tune batch size and epochs.

5. Model Evaluation -

- Accuracy score - Measures the proportion of correct predictions over the total predictions.
-

Confusion Matrix -
- True Positive (TP) - Correctly predicted churn.
- True negative (TN) - Correctly predicted non churn
- False positive (FP) - Predicted churn but customer stayed.
- False negative (FN) - Predicted stay but customer churned.

- Interpretation -
  - High TP and TN values indicate better model performance
  - FP and FN values highlight areas where the model makes mistakes.

CONCLUSION -

- Neural networks can effectively predict customer churn by learning complex patterns from multiple features.
- Preprocessing steps like normalization and categorical encoding are crucial for model performance.
- Model tuning (layers, neurons, learning rate, epochs) significantly improves accuracy.
- Evaluation using confusion matrix helps identify strengths and weaknesses in prediction.
- The classifier can assist banks in proactively identifying high risk customers and implementing retention strategies, thus reducing them.