National Vocational & Technical Training Commission
Institute of Electrical, Electronics and Computer Engineering
University of the Punjab

Artificial Intelligence C1 & C2
2nd Monthly Exam

Note: Plagiarism is a serious violation. Zero marks will be awarded in case plagiarism is found. Don't Copy Paste the Libraries And Your code

Note: Attempt all the Questions

Q1: SVM Model for Heart Disease Prediction

Scenario: You are a data scientist working with a medical research team. They have provided you with a dataset containing various features related to patients' health and whether they have heart disease or not. Your task is to build a Support Vector Machine (SVM) model to predict heart disease with at least 90% accuracy. Additionally, you need to compare the SVM model's performance with Decision Tree and Random Forest classifiers, and plot the feature with the highest variation.

- Predicting Heart Disease with SVM: You will start by preprocessing the data, splitting it into training and testing sets, and then training an SVM classifier on the training set to predict heart disease. Your goal is to achieve an accuracy of more than 90% on the test set.
- Generating the Confusion Matrix with Heatmap: Once the SVM model is trained, you need to generate a confusion matrix on the test set and visualize it as a heatmap. This will allow you to assess the model's performance in terms of true positives, true negatives, false positives, and false negatives.
- Comparing Results with Decision Tree and Random Forest: Next, you will build a Decision Tree and a Random Forest classifier using the same dataset and evaluate their performance on the test set. You'll compare the accuracy of all three models to determine which one performs best for this specific task.
- Plotting the Feature with the Highest Variation: To gain insights into the data, you will analyze the variation of each feature and plot the one that exhibits the highest variation. This will help you understand which feature has the most significant impact on predicting heart disease.

Q2: Artificial Neural Network for Bank Customer Retainability Prediction

Scenario: You work as a data scientist for a banking institution. Your manager has assigned you the task of building an Artificial Neural Network (ANN) model to predict customer retainability. The dataset contains various customer features and a target label indicating whether the customer stayed with the bank or left. Your objective is to create an ANN architecture that maximizes the prediction performance. Additionally, you need to handle categorical features, implement one-hot encoding and feature scaling, choose an appropriate number of epochs, and compare different activation functions using a confusion matrix.

- Encoding Categorical Features: Before building the ANN model, you will preprocess the dataset and handle any categorical features present in it. You'll choose an appropriate encoding method to convert categorical data into numerical format.

- Implementing One-Hot Encoder: To handle categorical variables effectively, you will implement a one-hot encoder. This will ensure that the ANN model can work with categorical data efficiently.
- Implementing Feature Scaling: To improve the convergence and performance of the ANN, you will apply feature scaling to the numerical features in the dataset. This will help ensure that all features have similar scales.
- Designing Hidden Layers for Maximizing Prediction: Based on the complexity of the problem, you will design the architecture of the ANN model with appropriate hidden layers, activation functions, and output layer for predicting customer retainability with high accuracy.
- Choosing Minimum Number of Epochs: You need to determine the minimum number of epochs required for the ANN to converge and achieve satisfactory performance. This will avoid overfitting while ensuring that the model has sufficiently learned from the data.
- Comparing Activation Functions Using Confusion Matrix: After training the ANN with different activation functions, you will generate a confusion matrix to compare their performance. This will help you identify which activation function yields the best results for this specific retainability prediction task.