Machine Learning Laboratory - Assignment III

Dataset Description

The dataset used in this exercise is the Breast Cancer dataset, which contains features related to cell nuclei measurements from breast cancer cases. The goal is to classify the samples into malignant or benign categories using a Decision Tree Classifier implemented from scratch, without using built-in functions.

Questions

1. Write Python code to perform an exploratory data analysis (EDA) on the dataset, which includes visualizing the distribution of key features.
2. Implement a Decision Tree Classifier from scratch using the Gini index as the criterion for splitting the dataset at each node.
3. Describe the recursive process for splitting the dataset into nodes based on the Gini index. Include conditions for stopping the recursion, such as the maximum depth of the tree or minimum number of samples per node.
4. Evaluate the performance of your Decision Tree Classifier on a test set. Report performance metrics such as accuracy, precision, recall, and the confusion matrix.
5. Visualize the confusion matrix and ROC curve.