Computer Science and Engineering Department, SVNIT, Surat

M.tech.-I, Semester-I

Machine learning (CSCS111) 2024-2025

Lab Assignment 4

Problem statement: Implement and evaluate methods for mitigating overfitting and underfitting during the training of machine learning models. Specifically, the task is to: Identify and address overfitting using techniques such as regularization (e.g., L1, L2), pruning, dropout (for neural networks), and cross-validation.

Evaluate the models on both training and test datasets to ensure they achieve a balance between fitting the training data and generalizing well to new data.

Success Criteria: The solution should be able to produce models that avoid both overfitting and underfitting, leading to improved performance on unseen data as measured by appropriate evaluation metrics (e.g., accuracy, F1-score, or RMSE depending on the task).

Note: For this practical, a Boston housing dataset will be used. This dataset includes information about housing prices in the Boston area, with 13 features describing aspects like crime rate, number of rooms, and proximity to employment centers.

Link: Available in Scikit-learn datasets: from sklearn.datasets import load_boston