

Machine Learning Course 2024 Spring: Experiment 4

May 19, 2024

1 Overview

In this experiment, you will implement the Decision Tree and two ensemble methods to combine multiple decision trees. You are expected to get insights into the tree structure and hyperparameters of the models.

2 Dataset Description

Covertypes dataset is a dataset about predicting forest cover type from cartographic variables only. The dataset is ready for you and is stored in the file '*Covertypes_train.csv*' and '*Covertypes_test.csv*'. Each row in the file is a sample, with the last column being the label and the first 54 columns being the features.

3 Task Description

3.1 Task 1

- Implement the **Decision Tree** algorithm.
- Visualize the tree structure.

3.2 Task 2

- Implement the **Random Forest** algorithm.
- Tune two hyperparameters, i.e. number of base learners and max depth of trees.

- Observe the influence of hyperparameters changes on accuracy.

3.3 Task 3

- Implement the **AdaBoost** algorithm.
- Tune two hyperparameters, i.e. number of base learners and max depth of trees.
- Observe the influence of hyperparameters changes on accuracy.

4 Requirements

- Implement learning algorithms with **Python**.
- Record the procedure of experiments in **Jupyter Notebook**. It is recommended to use **Markdown** in Notebook to properly explain the procedure.
- Your submission includes:
 - source code: python notebook (.ipynb);
 - ‘report.pdf’ file : your report. Refer to the report template for the format.