15.S60: Computing for Optimization and Statistics

Assignment 1: Machine Learning in Python

Due Date: Tuesday 16 of January, 11:59pm, 2024

January 10, 2024

1 Introduction

For this assignment, you will explore the tools learned in class to tackle a Machine Learning (ML) problem of your choice!

2 Assignment

Go to Kaggle and create an account or sign up if you are already familiar with the platform. Select a completed competition (either a regression or a classification task) of your preference. Create and submit (**both in Kaggle and in Canvas**) a notebook that has the following structure:

1. Start with the Exploratory Data Analysis:
    - Name 3 particular things about the data including plots. Be creative!
2. Run 2-3 models taught in class and report their performance (in sample and out of sample), using hyperparameter tuning.
3. Create a weighted ensemble using some of the models.
4. Submit your predictions to Kaggle and attach a screenshot of your public ranking.

In your final submission, you should have both the notebook you created (as pdf), as well as the screenshot of the kaggle leaderboard.

Some suggestions of public data:

- House Prices - Advanced Regression Techniques.
- Titanic - Machine Learning from Disaster.

For any questions, you can mail bbarri@mit.edu.