

15.S60: Computing for Optimization and Statistics

Assignment 3: Machine Learning in Python

Due: Tuesday, January 21st, 2025 at 11:59pm EST

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

Please visit [Kaggle.com](https://www.kaggle.com) and 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 some Data Preprocessing and Exploratory Data Analysis (EDA). Explain what your EDA plots show.
- 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 a PDF file), as well as the screenshot of the Kaggle leaderboard.

Some suggestions for Kaggle:

1. Regression Problem: [House Prices - Advanced Regression Techniques](#)
2. Classification Problem: [Titanic - Machine Learning from Disaster](#)

Have fun!

For any questions, email joyceluo@mit.edu.