

MSDS 630
HW 5
Gradient Boosting
Due: Mar 3, 2025 @ 11:59p

1 Gradient Boosting with MSE (15 pts)

1.1 Instructions:

Implement gradient boosting for the data found in “rent-ideal.csv”. In this exercise we’re trying to predict rent prices.

1. Write a program implementing gradient boosting using MSE as the loss. Use the template given in `gradient_boosting_mse.py` as starter code.

Some functions are pre-written according to their descriptions, others need to be written by you. These functions are indicated with sections like:

```
### BEGIN SOLUTION
```

```
### END SOLUTION
```

The specific directions for each function are found in the function docstring.

- Do not submit full data to Github. Doing so will result in a loss of 5 points. Only commit the tiny datafiles.

2. Set the learning rate to 0.1. Apply gradient boosting to your dataset using different values for `numTrees`. How do you find the best value for `numTrees`? Report the train and validation R^2 for the best value of `numTrees`. Make a figure which summarizes your experiment by plotting training and validation R^2 as a function of the number of trees. Try at least 2000 trees.
3. Compare your results with the results of running the gradient boosting package XGBoost. Explore the hyperparameters given in the package. Make plots or tables which illustrate your tuning process and the results of your experiments.

1.2 Deliverables:

Submit your completed functions as a Python script entitled “`gradient_boosting_mse.py`” along with the write ups and plots of your experiments for parts 2. and 3.

1.3 Evaluation:

Once you've finished, make sure you can run:

```
pytest test_gradient_boosting_mse.py
```

There are 3 tests to pass. Each pass is worth 5 points. If you fail an assertion, but your code is on the right track, half credit will be awarded for that check.