Data Ventures: Practical 1

Predicting House Prices

Spring 2018

1 Description

Your goal is to train a model to predict house prices as accurately as possible. You'll receive a csv with approximately 10,000 rows, each with 22 attributes (columns) including a price. You'll receive another csv with approximately 10,000 rows with the price missing. Fill in the 10,000 missing prices as accurately as possible.

2 Rules

Outside packages are allowed, but outside data isn't! No pretrained models. As you make submissions, the leaderboard will be updated. The leaderboard is only an estimate of your true score. Half of the test data will be set aside for the final score calculations (to avoid overfitting on the testing set, and incentivizing a large number of submissions).

Scoring will be done based on average R^2 accuracy.

3 Baseline

In order to receive credit for the assignment you must beat the baseline, an off the shelf Random Forest model, with a R^2 of 0.84. The code for this is on the github.

4 Submission

The upload format is a CSV with a 1 column with the predicted price.

Click here: `https://goo.gl/forms/PDGqATnB3obWboxR2`