Individual Homework

Due: 11.59pm on Friday, February 9

Homework instructions:

- This is an individual homework assignment. It consists of one prediction problem, similar to what you were working on in the group homework 2. You can discuss with other people, but you need to create your own writeup and your own code for submission. Also keep in mind that your score will be relative to other student's submissions.
- Submit your write-up to gradescope.com. In addition, you will email your predictions to boothmlteam@gmail.com.

Files needed for this homework can be downloaded here:
https://github.com/ChicagoBoothML/ML2017/tree/master/hwind

Question

You are provided with a data set obtained from the Ames Assessor's Office and is used for tax assessment purposes. In this homework, you will use it to predict home selling prices. The type of information contained in the data is similar to what a typical home buyer would want to know before making a purchase.

Description of the variables can be found in the file `description.txt`. The data contains 23 nominal, 23 ordinal, 14 discrete, and 20 continuous variables. Parcel identification number (PID) can be used with city web site for parcel review, though this is not needed for the assignment (See http://www.cityofames.org/assessor/ and click on "property search").

You are provided with two files: `HousingTrain.cvs` contains the training set with 1925 observations and `HousingTest.cvs` contains test data with 1000 observations that we will use to evaluate your model. Your task is to build an accurate model for predicting `SalePrice`.

Predictions will be evaluated using the root mean squared error (RMSE) on the logarithm of the predicted value and the logarithm of the observed sales price

$$\sqrt{\frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \left(\log(\text{SalePrice}_i) - \log(\widehat{\text{SalePrice}_i}) \right)^2}.$$

Taking logs means that errors in predicting expensive houses and cheap houses will affect the result equally.

Save your predicted `SalePrice` in a file `hwind-<your_uchicago_id>.csv`, where you will need to replace `your_uchicago_id` by your UChicago ID. Your file should contain only one column with a header `SalePrice` and 1000 entries of predicted values.

You should email your submission file to boothmlteam@gmail.com and another file with the code you used to make predictions.

You will receive points based on your write-up, whether we can compute RMSE based on your submission and your relative ranking in the class.