Zillow's Home Value Prediction (Zestimate)

Project overview

Zillow.com home valuation gives consumers as much information as possible about homes and the housing market, marking the first time consumers had access to this type of home value information at no cost.

"Zestimates" are estimated home values based on 7.5 million statistical and machine learning models that analyze hundreds of data points on each property. Zillow has continually improved the median margin of error (from 14% at the onset to 5% today).

The objective of this project is to create a model to predict either:
1. log error of Zestimate (regression), **or**
2. sign of log error (classification)

Log error is defined as:

$$logerror = log(Zestimate) - log(Sale\ Price)$$

Note that predicting "sign of log error" (classification) is added to simplify the project for this course and also teach important classification concepts.

Necessary background for success

It is expected that all students have some experience with Python and have some familiarity with basic statistical and linear algebraic concepts such as mean, median, mode, standard deviation, correlation, and the difference between a vector and a matrix. In Python, it will be helpful to know basic data structures such as lists, tuples, and dictionaries, and what distinguishes them (that is, when they should be used).

Environment/Software Requirements

We will be using Anaconda with Python 2.7. Note that we'll use Python 2.7 not 3.6. Anaconda comes with the following tools and libraries which we'll use:
- Jupyter Notebook (a.k.a. iPython Notebook)
- Scikit Learn
- Pandas
- Numpy, Scipy
- Matplotlib

Hardware requirements:

- Operating system: Windows XP or newer, 64-bit macOS 10.9+, or Linux, including Ubuntu, RedHat, CentOS 6+, and others.
- 5 GB disk space for Anaconda and data