Zillow's Home Value Prediction

1 Project Idea

The idea of the project is to give a better estimate of a house as it is the largest and most expensive purchase a person makes in his or her lifetime. Zestimate (predicted price) was created by Zillow, to give 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. Different aspects of a house such as square feet area, location, number of rooms, tax value and many more, play an important role in estimating the house price.

Zillow provides full list of real estate properties in three counties (Los Angeles, Orange and Ventura, California) from 2016. Here we try to predict the log-error between their Zestimate and the actual sale price, given all the features of a home and minimize it. We will start with feature extraction or reduction and then move on to implementing different machine learning techniques for regression. We will then select the best method which gives the lowest log error.

2 Dataset

The data set consists of a data dictionary which provides about 60 attributes relevant for evaluating the house price. The train data has all the transactions before October 15, 2016, plus some of the transactions after October 15, 2016 (90,000). The test data has the rest of the transactions between October 15 and December 31, 2016.
train.csv - training data , test.csv - test data

3 Software

We would implement the project in either Python or R and test the results. For Machine Learning Algorithms , API's from scikit learn and numpy will be used. We would be using VCL resources for running our programs.

4 Midterm Milestone

Present data exploration techniques.Select feature selection or extraction techniques to reduce attributes such that they represent maximum information about the data-set.Implement one regression algorithm to predict the house price.

5 Work Division

Harish Pullagurla: Data cleaning, **Ratika Kapoor**: Feature reduction/extraction ,
Pooja Mehta: Prediction technique

6 References

[1] Park.B & Jae.K, Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data,(2016) *Expert Systems with Applications*

[2] Xibin.W ,Junhao.W, Yihao.Z & Yubiao.W Real estate price forecasting based on SVM optimized by PSO. (2014) *Optik - International Journal for Light and Electron Optics*

[3] Gu.J, Zhu.M, & Jiang.L, Housing price forecasting based on genetic algorithm and support vector machine,(2011)*Expert Systems with Applications*