Project Proposal

House Price Prediction

[1]. Wang Tianqi 1401213465
[2]. Matta Uma Maheswara Reddy 1601213442

Project Description

This is a Kaggle Competition to predict housing sale prices using advanced regression techniques. There are 79 variables that can be used for predicting the housing prices. We plan to implement various techniques we have learned in the class and try to predict the housing prices.

Project Goal

Our job is to predict the sales price for each house, which is stored in 'SalePrice' column, and try to minimize the Root-Mean-Squared-Error (RMSE), which equals to:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n}}$$

Dataset

All the data are provided from Kaggle, which has already been separated into train dataset (with 1460 observations) and test data (with 1459 observations). There are 79 features.

Methodology

- Since there are plenty features, we expect to spend most of time on feature engineering.
- Since the project is a regression problem, we would try some advanced regression techniques based on algorithms, e.g decision tree regressor, random forest regressor.