Machine Learning Project Proposal

Xicheng XIA

Jun 9th 2019

1 Project description

We decided to choose *House Prices: Advanced Regression Techniques* from the Kaggle competition as our project topic. Our aim is to use simple machine learning models such as *Lasso Regression* to obtain a decent result after applying feature engineering technique.

2 Data Description

This competition requires us to predict the exact house price with 79 original features. Train set is 1459×79 with given house prices for training, test set is 1459×79 without knowing house prices.

3 Methods of Research

a Read those kernels;

b Data cleansering. Cope with missing values and abnormal values(median filter, winsorize);

c Single factor exploration;

d Feature engineering. Dummies, linear/non-linear transformations;

e Dimention reduction. PCA;

f Scale the data. Normalization/Standarization, max_min_scaler.

g Select models and parameters. Lasso/Ridge/SVR, cross-examination/regularization/penalization/learning-rate/iteration-times

4 Our Goal

The competition score is measured by the out-of-sample *Root Mean Squared Logarithmic Error*, we expected to at least obatin a score less than 0.15 (ranked around 3000), an optimistic expectation is 0.13 (ranked around 2000).