Zillow's Price Training

Chan Yu Yankai Liu Yifei Bi

Procedures Guide

01

02

1.XGBoost
2.LightGBM
3.Neural Network
4.Linear Regression

03

04

Reporter : Chan Yu

1

XGBoost - Extreme Gradient Boosting

- Regularization
- Parallel Processing
- High Flexibility
- Handling Missing Values
- Tree Pruning

1.XGBoost
2.LightGBM
3.Neural Network
4.Linear Regression

XGBoost - First

Combined XGBoost Predictions

XGBoost - First

Set XGBoost Hyper Parameters

```
xgb_params = {
    'eta': 0.037,
    'max_depth': 5,
    'subsample': 0.80,
    'objective': 'reg:linear',
    'eval_metric': 'mae',
    'lambda': 0.8,
    'alpha': 0.4,
    'base_score': y_mean,
    'silent': 1
}
num_boost_rounds = 250
```

```
First XGBoost predictions:
          0
0  0.003354
1  0.008021
2 -0.002644
3  0.009713
4 -0.001039
```

Validate For First XGBoost:

0.06695654820228315

Combined XGBoost Predictions

Combined XGBoost predictions:

	0
0	0.003386
1	0.008179
2	-0.002297
3	0.009634
4	-0.000765

XGB1_WEIGHT = 0.8000
XGB2_WEIGHT = 0.2000

Validate For Combined XGBoost:

0.06694858812380969

2

LightGBM

1.XGBoost
2.LightGBM
3.Neural Network
4.Linear Regression

- Optimization in Speed and Memory Usage
- Optimization in Accuracy
- Optimization in Network Communication
- Optimization in Parallel Learning

```
Unadjusted LightGBM predictions:
          0
0  0.021860
1  0.022435
2  0.022740
3  0.021015
4  0.022069
```

Validate For LightGBM

0.06942165970198744

3

Fitting neural network model...

0.0699696842884

Neural Network predictions:

	parcelid	201607	201608	201609	201707	201708	201709
782	14677191	0.004404	0.004402	0.004400	0.004405	0.004403	0.004401
968	11183209	0.004399	0.004403	0.004407	0.004400	0.004403	0.004407
1165	11554091	0.004373	0.004375	0.004378	0.004376	0.004377	0.004379
1351	11742566	0.004386	0.004388	0.004393	0.004387	0.004389	0.004394
1609	14667297	0.004405	0.004404	0.004403	0.004406	0.004405	0.004404

Validate For NN

0.06760387642306702

1.XGBoost
2.LightGBM
3.Neural Network
4.Linear Regression

Model Structure

Input
400 hidden units
relu
dropout 0.4
160 hidden units
relu
dropout 0.6
64 hidden units
relu
dropout 0.5
26 hidden units
relu
dropout 0.6
1 hidden units

4

Linear Regression

1.XGBoost
2.LightGBM
3.Neural Network
4.Linear Regression

OLS predictions:

	parcelid	201607	201608	201609	201707	201708	201709
782	14677191	-0.006553	-0.008514	-0.010476	-0.002206	-0.004168	-0.006130
968	11183209	0.013480	0.011519	0.009557	0.017826	0.015865	0.013903
1165	11554091	-0.005324	-0.007286	-0.009248	-0.000978	-0.002940	-0.004901
1351	11742566	0.023170	0.021208	0.019246	0.027516	0.025554	0.023593
1609	14667297	-0.017631	-0.019592	-0.021554	-0.013284	-0.015246	-0.017208

Validate For OLS

0.0680084796120211

Combination

- Combining XGBoost, LightGBM, and baseline predictions...
- XGB_WEIGHT = 0.6200
- BASELINE_WEIGHT = 0.0100
- OLS_WEIGHT = 0.0620
- NN_WEIGHT = 0.0800

```
Combined XGB/LGB/baseline predictions:
          0
0  0.007198
1  0.010301
2  0.003876
3  0.010879
4  0.004672
```

Combination

- Combining with XGB/LGB/NN/OLS/baseline predictions…
- XGB_WEIGHT = 0.6200
- BASELINE_WEIGHT = 0.0100
- OLS_WEIGHT = 0.0620
- NN_WEIGHT = 0.0800

Combined XGB/LGB/NN/baseline/OLS predictions:

	parcelid	201607	201608	201609	201707	201708	201709
782	14677191	0.0080	0.0079	0.0077	0.0083	0.0082	0.0080
968	11183209	0.0129	0.0127	0.0126	0.0132	0.0130	0.0129
1165	11554091	0.0044	0.0042	0.0041	0.0047	0.0045	0.0044
1351	11742566	0.0142	0.0141	0.0139	0.0145	0.0144	0.0142
1609	14667297	0.0044	0.0043	0.0041	0.0047	0.0046	0.0044