Machine Learning

Sberbank Russian Housing Market

TAs
ntu.mlta@gmail.com

Task description

- In this challenge, complex interactions between housing features such as number of bedrooms and location are used to make pricing predictions .

https://www.kaggle.com/c/sberbank-russian-housing-market

https://www.kaggle.com/c/sberbank-russian-housing-market

Data Files

- **train.csv, test.csv** : information about individual transactions.
- **macro.csv** : data on Russia's macroeconomy and financial sector
- **data_dictionary.txt** : explanations of the fields available in the other data files

Taining & Testing data

- 292 attributes per transcation
- continuous & discrete attributes
- including information of a house and local area.
- you can see explanation in data_dictionary.txt
- macro.csv could be joined by timestamp

Evaluation - RMSLE

$$\epsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i + 1) - \log(a_i + 1))^2}$$

n is the total number of observations in the (public/private) data set,

p_i is your prediction,

a_i is the actual response for i.

$\log(x)$ is the natural logarithm of x

Evaluation - RMSLE

$$\epsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i + 1) - \log(a_i + 1))^2}$$

n is the total number of observations in the (public/private) data set,

p_i is your prediction,

a_i is the actual response for i.

$\log(x)$ is the natural logarithm of x

- RMSLE penalizes an under-predicted estimate greater than an over-predicted estimate.

Characteristics of this task

- More attributes and instances of data
- Economic prior knowledge
- Prize money!!!!!!!!!!!!!

Reminder

- Some missing labels
- Seperate discrete and continous attributes first.