Data Mining Prediction Assignment

10 January, 2017

File descriptions

- **train.csv**: the training set
- **test.csv**: the test set
- **linear_model_prediction.csv**: a sample submission file in the correct format
- **Linear_model_example.Rmd**: R code for producing the linear model predictions.

Data fields

- **"price"**: price of the house. (indeed, this field is not present in "test.csv" since it needs to be predicted).
- **"x" and "y"**: coordinate of the house.
- **"bedrooms" and "bathrooms"**: number of bedrooms and bathrooms.
- **"land_size" and "living_area" and "basement"**: area of the property, of the living space, and of the basement (if any).
- **"floors"**: number of floors.
- **"quality"**: quality of the built of the house.
- **"environment"**: quantifies the quality of the surroundings of the house.
- **"maintenance"**: quantifies how well the house has been maintained.
- **"year_built" and "year_renovation"**: year when the house was build, and renovated. If year_renovation=0, the house has never been renovated.
- **"Id"**: index id of the house

Evaluation

Submissions will be evaluated using th Root Mean Square Percentage Error (RMSPE). If $\hat{p}_i$ is your estimate of the (true and unkown to you) price p_i of the i-th house, then your final score is:

$$(\text{RMSPE}) \;=\; \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{p}_i - p_i}{p_i}\right)^2}$$

Rules

- You can use any approach and programming language you deem appropriate. You are indeed encouraged to read research papers, books, and implement many different ideas.
- Note that you will need to submit a report describing your approach and that your code should be reproducible (i.e. it is a good idea to keep your code clean from the very start of your investigations). If it is not possible to reproduce your predictions, your submission will not be validated.
- You are **not** allowed to use any other data-set except the one provided in this competition in order to make predictions.

First price

Full mark ($\approx 5\%$ of total) and fame!