By Kyrylo Pogrebenko

Predicting loan rates.

Attempting to solve the problem of predicting loan rates I used two algorithms: RandomForest, ElasticNet. These two differ fundamentally in a way of identifying the best hyper parameters. RandomForest is an ensemble learning method multiple Decision Tree algorithms which in a case of regression produces the mean of individual trees. Where the ElasticNet algorithm is a complex approach to a classical Linear Regression with regularized regression method of assessing the penalty for overfitting, L1 Lasso or L2 Ridge methods.

Before applying machine learning algorithms, I massaged the raw data provided, removing one row with all NULLs, but the target cell X1. For two types of categorical variables, I one-hot encoded nominal features and labeled ordinal features, expanding the total number of features to train on. Continuous features needed to be transformed and normalized in order not to inflate the overall variance. Kindly find corresponded plots of all continuous features in "plots" folder, which show density plots of attempts of transformation: x^2, x^(1/4), ln, log2, log10, ln(1+p), and BoxCox. Included file addons.py contains comments with explanations which method was applied for given features.

Note:

Please keep in mind that I modified the input files, by just only changing their names to self-explanatory ones. Also I 0'ed them to be able to email a project folder.

Solving broken code.

Task1. Incorrect data type was used in print statement.

Task2. Count_a++ was replaced with count_a += 1

Task3. Positional parameters were adjusted, as well as expanding the map-list.

Task4. Sorted built-in function for lists.

Task5. Multiplication added in the recursive step of the function itself.