

Springleaf

Guschin Alexander
fall 2015

Determine whether to send a direct mail piece to a customer

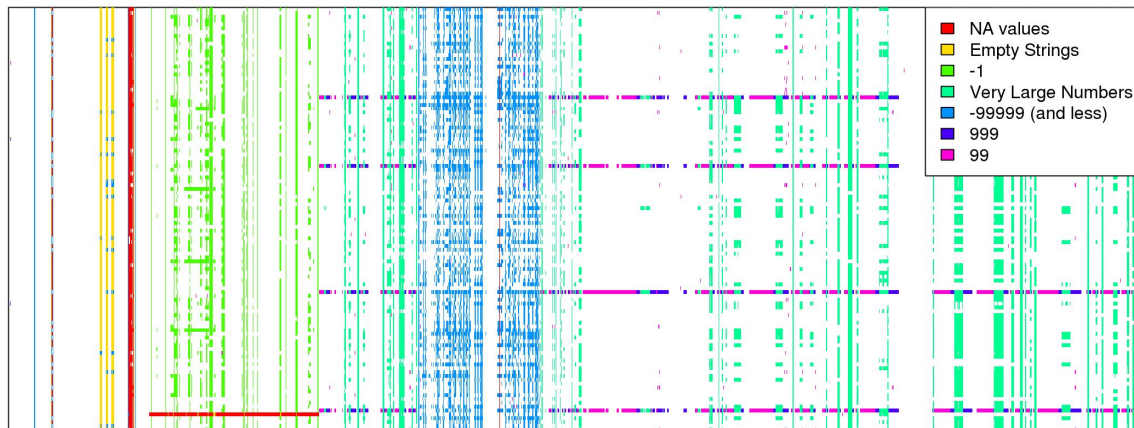
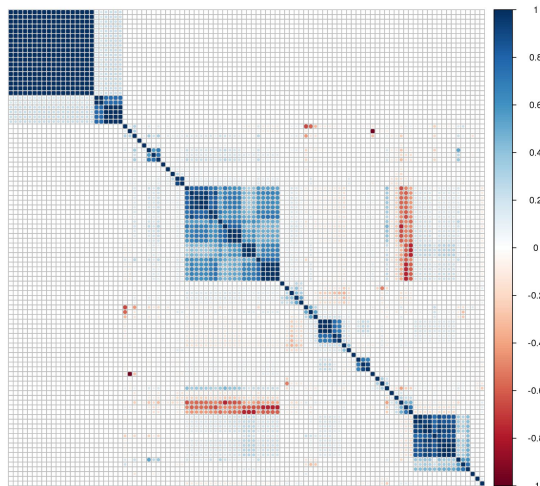
145k objects in train

2 classes

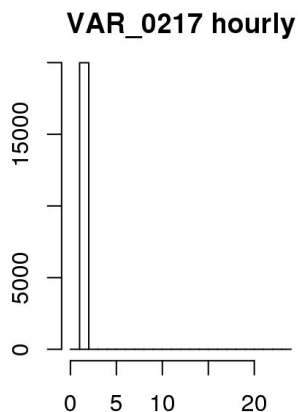
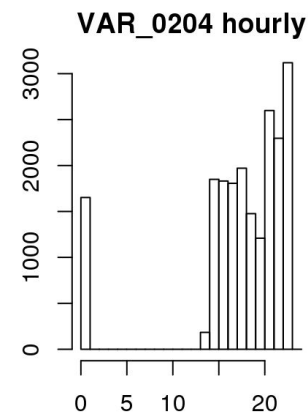
anonymized features (~2k)

AUC

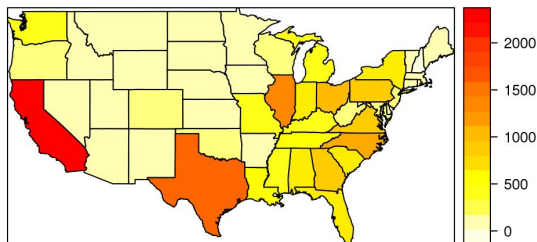
Pearson correlation of 100 sampled numerical features



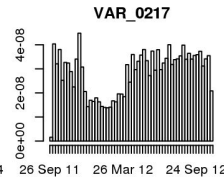
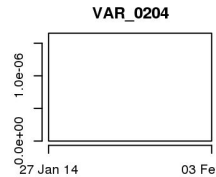
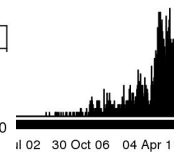
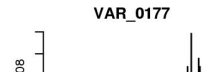
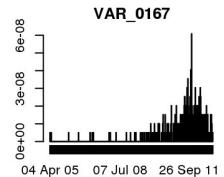
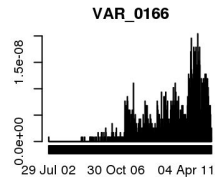
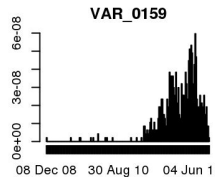
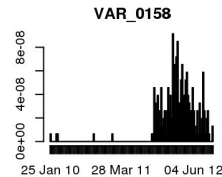
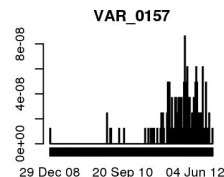
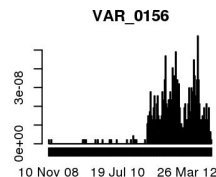
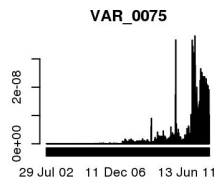
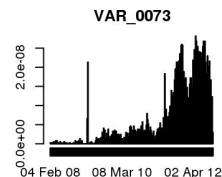
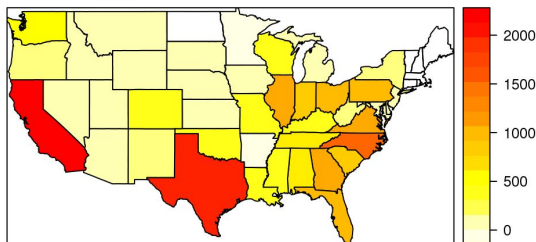
Hope for smart feature engineering



VAR_0274 value count



VAR_0237 value count



feature packs

0 level

processed
dataset

basic dataset - Basic data cleaning and feature engineering

basic dataset

processed dataset - Advanced data cleaning and feature engineering

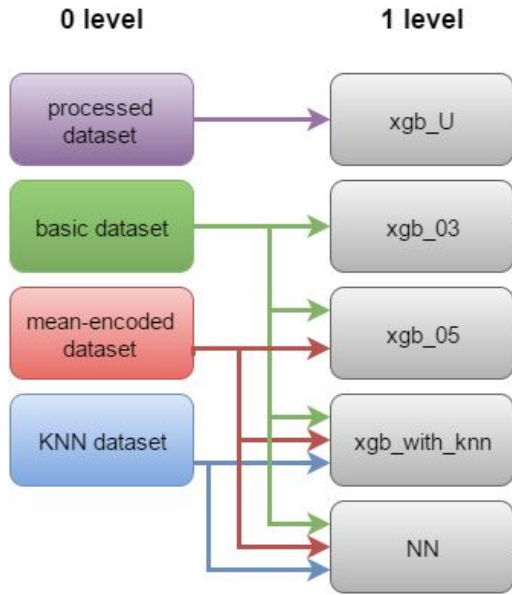
mean-encoded
dataset

mean-encoded dataset - Projecting features into homogeneous space

KNN dataset

KNN dataset - distance features on mean-encoded

1 level : out-of-fold predictions - xgboost

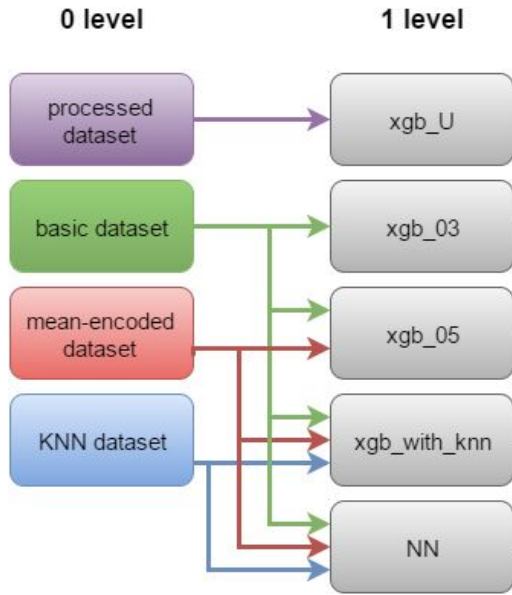


oof predictions (== metafeatures) should be diverse

main idea: each metafeature should bring 'new' information about Y

Xgboost trick: $\eta * \text{num_rounds} \approx \text{const}$

1 level : out-of-fold predictions - NN



ranks:

$\text{rank}([0.1, 4, -2]) = [2, 3, 1]$

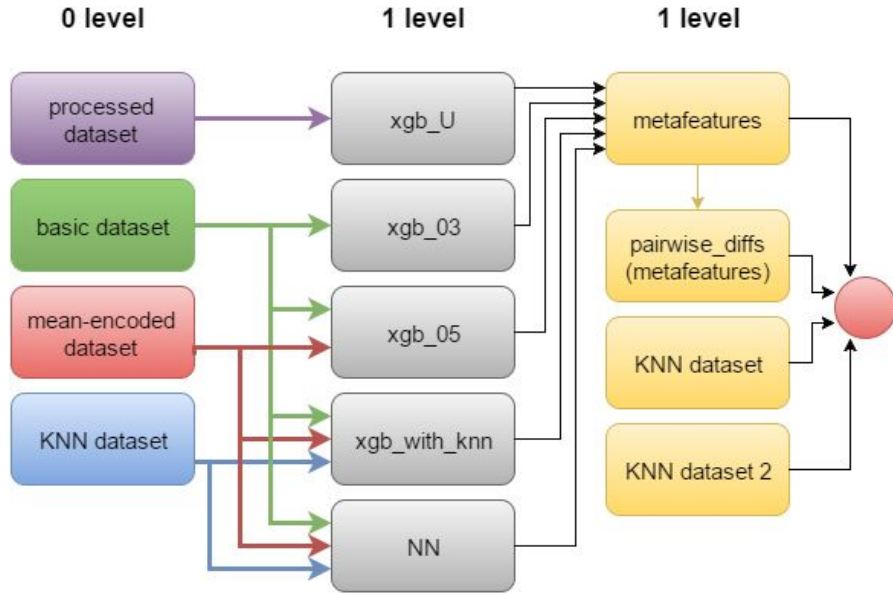
StandartScaler:

$x = (x - \text{mean}(x)) / \text{std}(x - \text{mean}(x))$

power:

$\text{feature} = \text{sign}(\text{feature}) * |\text{feature}| ** 0.5$

1 level - features

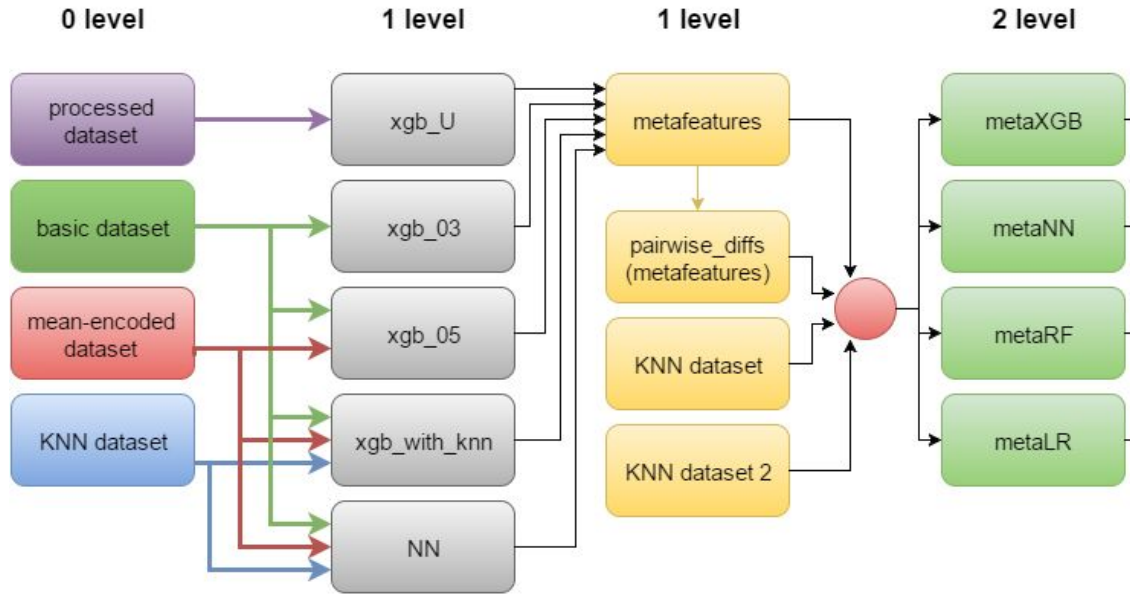


KNN dataset - distance features on mean-encoded

KNN dataset 2 - distance features on mean-encoded ** 0.5

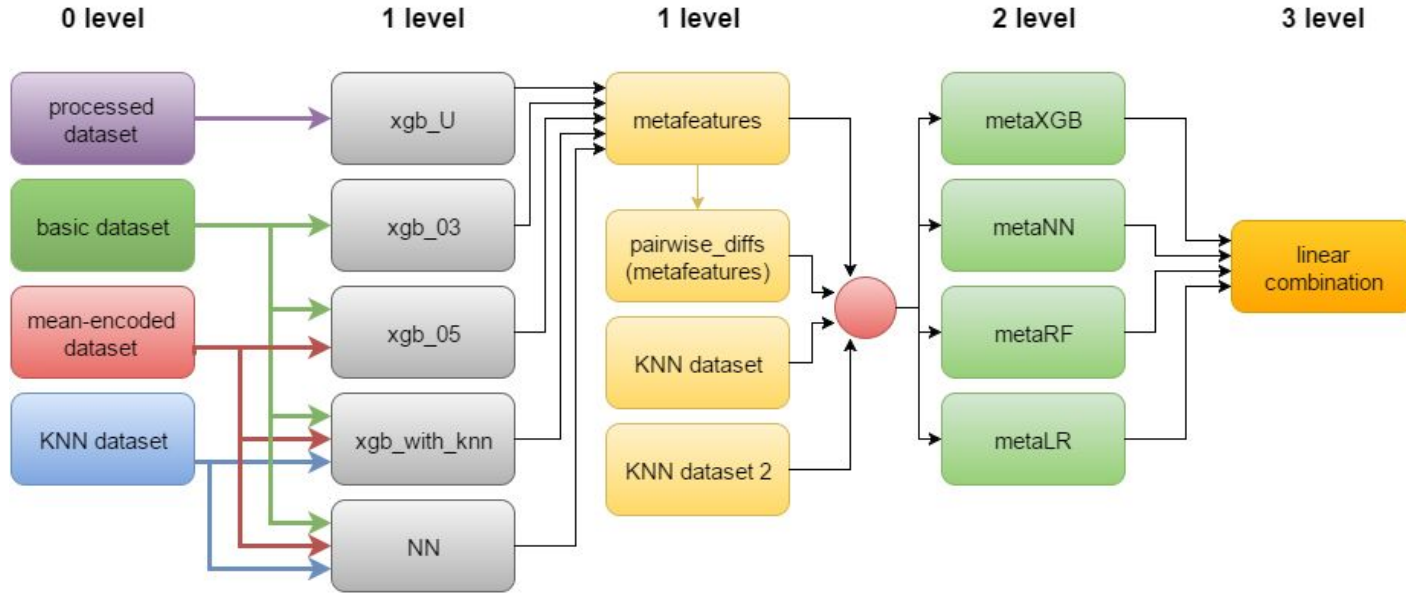
Additional features should bring new information to 1 level

2 level - classifiers



- classifiers should be simple
- predictions should be diverse

3 level - final



coefficients
are
estimated
using oof

scripts on kaggle:

<https://www.kaggle.com/steves/springleaf-marketing-response/visualizing-na-values/notebook>

<https://www.kaggle.com/darraghdog/springleaf-marketing-response/explore-springleaf>