1. data explore:
   1. classified variables into 4 by pattern richness and
2. month effect model: split 2016 10/11/12 month data randomly into 2 sets, train 2016 & 2017 rest of the data together with half of 10/11/12 data, predict on
   1. All data used in training.
   2. half of 10/11/12 data used in training;
   3. other half of 10/11/12 data as validation

And measure median error on each of the prediction target.

Number of samples for 10/11/12: 8542.

1. Train without sale\_month as feature:

IS predict miss median: -0.0000324

IS predict miss median on month: 0.0033405

OS predict miss median: 0.0058031

1. Train with sale\_month as feature:

factor is fairly used in prediction, avg feature importance rank 10.

IS predict miss median: -0.0000361

IS predict miss median on month: 0.0036768

OS predict miss median: 0.0053516

20171015:

1. Class3 feature engineering, use ‘best’ group mean to substitute original high cardinal class variables, successfully improved both IS cv and OS score.
2. Step3 prediction research, does not improve final error median when using 2016 month data to train model and predict 2017 month. It is reasonable, joined year 2016 -> 2017 training does not differentiate months, so this type of joined training only learns the ‘mean’ error pattern across months. Yet according to actual 2016 & 2017 error median, it is quite different (sign) across months. So cannot expect to improve each month’s result through joint training.
3. ‘best’ number of boosting rounds can be learned by length of cv\_hist of LightGBM.cv. Re-param-search for all models again. Lower learning rates leads to more stable model (for month-train, it means smaller best\_n\_rounds std.)