

Appendix

Comments:

1. After selecting a series of models and setting their parameters, we use the training data of first 33 months (date_block_num<33) to fit models and use the 34th historical sales data that is October 2015 (date_block_num=33) as the validation set. Eventually, based on the fitted models, we make prediction of November 2015's sales (date_block_num=34) and submit the test file to Kaggle for the RMSE score.

2. We have done the tuning process on a set of models, some failed due to computer hardware limitation (not enough memory). Nevertheless, we had tried and made every effort to gain better performance of the models and to do as best as we can in the project.

Table 1 Initial models and results

Models	Initial parameters	RMSE
Linear	copy_X: True, fit_intercept: True, normalize: False	1.01083
Ridge	'alpha': 128, 'copy_X': True, 'fit_intercept': True, 'normalize': False	1.01060
Lasso	'alpha': 128, 'copy_X': True, 'fit_intercept': True, 'normalize': False	1.01059
Bayesian ridge regression	'alpha_init': 1 'lambda_init': 1	1.01016
AdaBoosting	'base_estimator' : DecisionTreeRegressor(), 'base_estimator__min_samples_leaf': 5 'n_estimators': 50, 'learning_rate': 0.01	1.03695
XGBoost	Max_depth=10, n_estimators=1000, min_child_weight=0.5, colsample_bytree=0.8, subsample=0.8,eta=0.1,seed=42	0.91755
LightGBM	'objective': 'rmse', 'metric': 'rmse', 'num_leaves': 35,	0.91315

	'min_data_in_leaf':10, 'feature_fraction':0.5, 'learning_rate': 0.05, 'num_rounds': 900,	
ExtraTrees	n_estimators=50 random_state=0	0.94308
Random Forest	n_estimators=60, min_samples_leaf=15, n_jobs=-1, random_state=0	0.93246
Gradient Boosting	random_state=0	1.13437
LSTM	epochs = 20, batch = 128, lr = 0.0001, Activation: Relu, Optimizer: Adam	1.01024

Table 2 Tuned models and results

Models	Tuned parameters	RMSE
Bayesian ridge regression	'alpha_init': 1.5, 'lambda_init': 1e-07	1.01016
AdaBoosting	Tuning failed due to memory limit	
XGBoost	max_depth=8, n_estimators=1000, min_child_weight=300, colsample_bytree=0.8, subsample=0.8, eta=0.3, seed=42, num_leaves= 1023, min_data_in_leaf=10, feature_fraction=0.7, learning_rate= 0.01, early_stopping_rounds = 30, gamma = 0.8	0.91616

LightGBM	Objective : 'rmse', metric: 'rmse', num_leaves: 1023, min_data_in_leaf:10, feature_fraction:0.7, learning_rate: 0.01, #0.01 num_rounds: 1000, early_stopping_rounds: 30, seed: 1	0.90984
ExtraTrees	random_state=10, max_depth=50, min_samples_split=10, min_samples_leaf=40, n_estimators=50	0.89930
LSTM (Technical issue, no GPU)	batch_size = 4096, epochs = 10, lr = 0.0001 Activation: Relu Optimizer: Adam	1.00609
RandomForest	Tuning failed due to memory limit	