FRE 9733 Big Data in Finance Week 10 Homework

Mengheng Xue

April 30, 2019

1 Best Models for LR, RF and MLP

Based on previous assignments, my choices of best model settings are as follows

	Best LR model setting
sample size	100k
regressors	AGE, MTM_LTV, INCENTIVE, CREDIT_SCORE, ORIG_INTRATE, UNIT_COUNT
hyper-parameters	MaxIter = 200, RegParam = 0.002, ElasticNetParam = 0.008

	Best RF model setting
sample size	100k
regressors	AGE, MTM_LTV, INCENTIVE, CREDIT_SCORE, ORIG_INTRATE, UNIT_COUNT
hyper-parameters	NumTrees = 12, MaxDepth = 2

	Best MLP model setting
sample size	100k
activation function	AGE, MTM_LTV, INCENTIVE, CREDIT_SCORE, ORIG_INTRATE, UNIT_COUNT
hyper-parameters	solver = gd, hidden layers = $[6, 6, 6]$

2 Performance Evaluation

The following is the performance comparison based on the setting in previous section. We can see that:

- LR and RF models beat the vacuous model on both clean and dirty dataset.
- MLP does not beat vacuous model but it obtains the best performace on dirty dataset.
- The performance on dirty dateset for all three models outperform performance on clean dataset, which is counterintuitive.

label	avg(distEntropy)	avg(rho)
vacuous_pred	0.20681209714500048	0.19089326539426685
rf_pred	0.2054605528719031	0.18670712257437866
lr_pred	0.20582348600876407	0.18664403996782625
mlp_pred	0.21624981463894227	0.19367240736147845
rf_pred_dirty	0.24291524989380764	0.15147378607250267
lr_pred_dirty	0.26086423150057797	0.15247867358668282
mlp_pred_dirty	0.20043043749170192	0.13845622551056694

Fig. 1: performance comparison

From following figure, we could see that for vacuous model, no group of loans from the vacuous model goes over 5.0.

round(rho, 0)	rhoSum	count
0	549.440314050421	11982
3	1477.8336151393548	440
5	358.89188823864725	78

Fig. 2: vacuous model