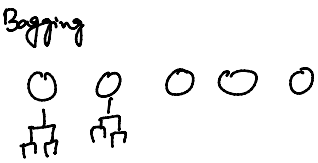
Boosting Cont...

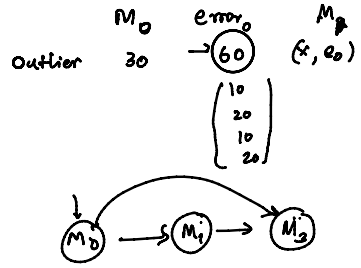
15 December 2022 20:52

Boosting — XGBoost, LightGBM



Disadvantages of Boosting (GBDT)

- → Sequential learning ✓ (Slow)
- → overfitting ✓
- → outliers significal Impact (GBT) ✓



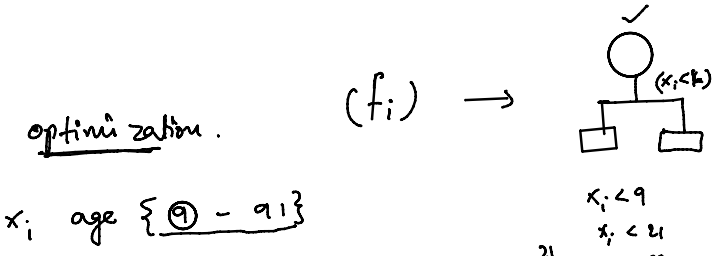
(Stochastic Gradient Boosting)

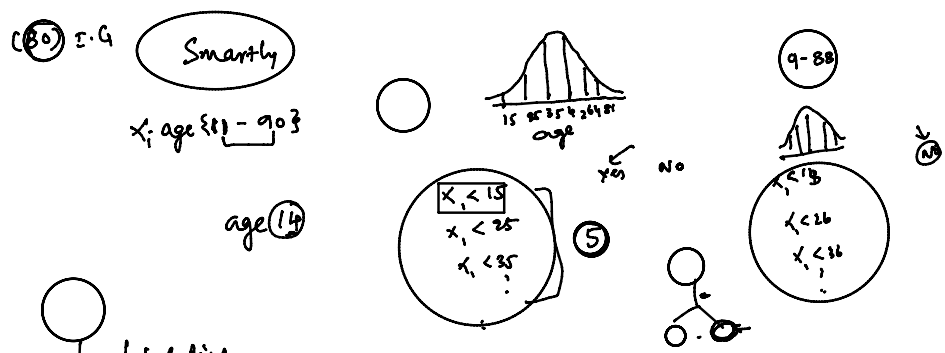
✓XGBoost — Row Sampling + Column Sampling

(RF) (diversify)

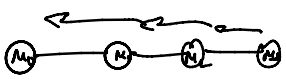
(+ optimization)

→ Feature level optimization.



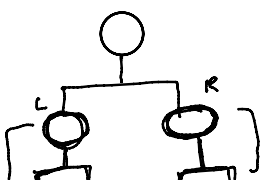


② Parallelization



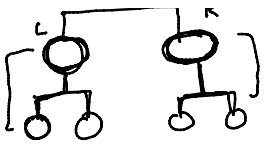
Xtreme Gradient Boosting

(2016)



(Parallization)

Xtreme Gradient ... (2016)



(Parallization)

③ Hyper Parameters (XG Boost) ← (our Stem)

n_estimators → # of Base models

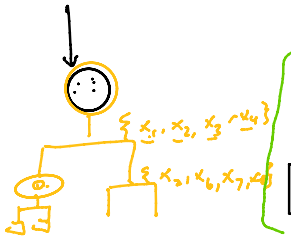
Eta → (Learning rate)

(learning rate)

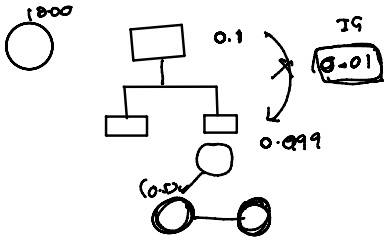
$M = \alpha_0 M_0 + (\gamma)\alpha_1 M_1$

Max_depth

min_split_loss = 0.05



Sub Sample : (0.5) [0.6 - 0.8]



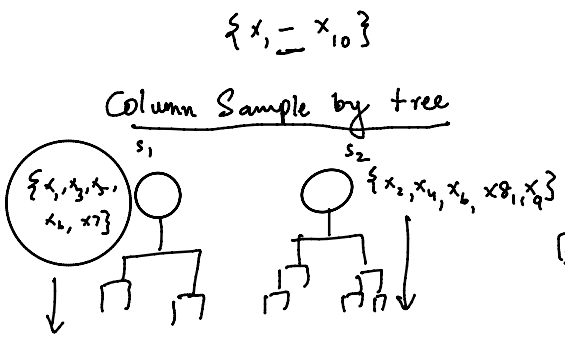
Col Sample by tree (=0.5), Col Sample by split, Col Sample by node

Regularization

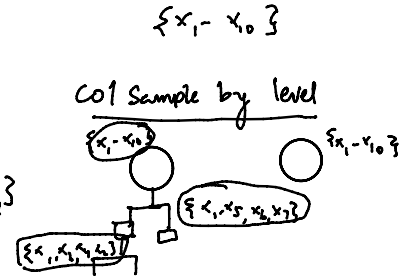
$M = \alpha_0 M_0 + \alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3$

y	$\hat{y}$
10	$(\alpha_0 M_0 + \alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3)$
20	
30	
40	

$$\sum_{i=1}^{n} \left(10 - (\alpha_0 M_0 + \alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3)\right)^2 + \lambda|\alpha_i|$$

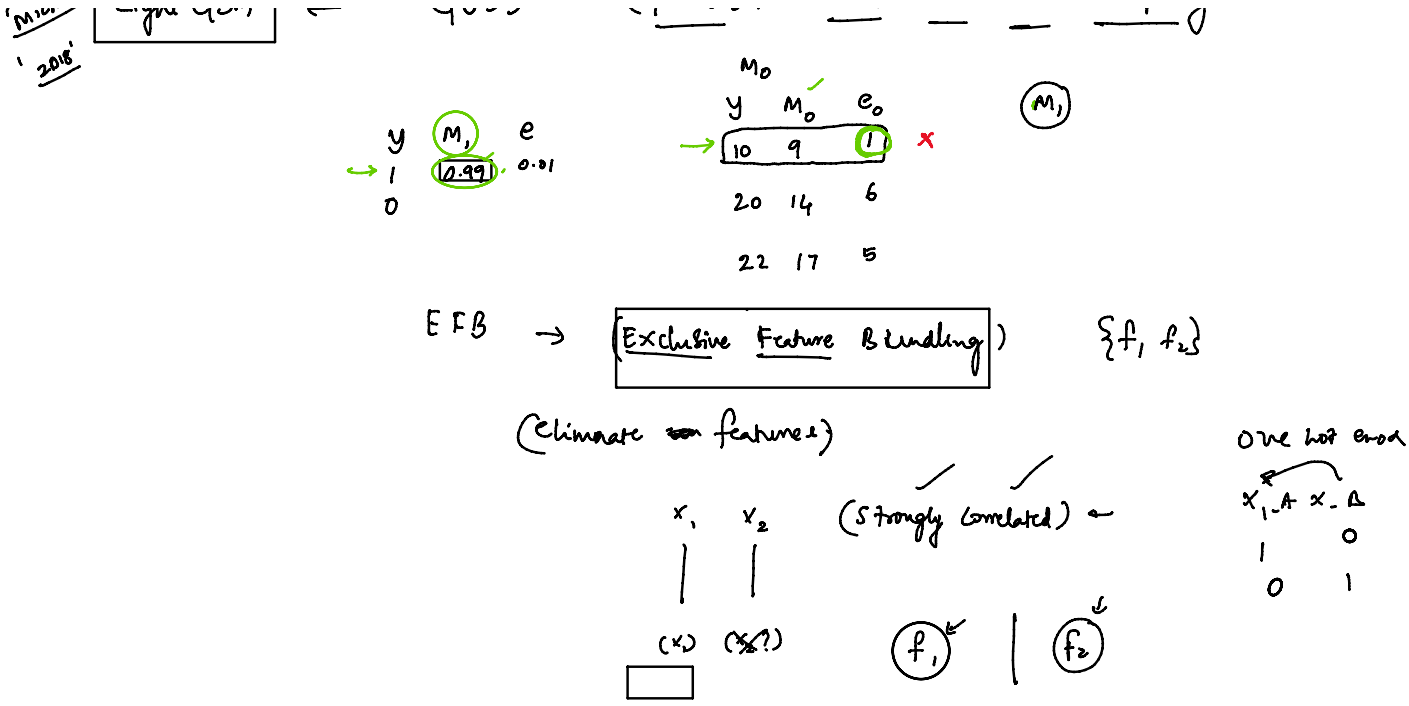


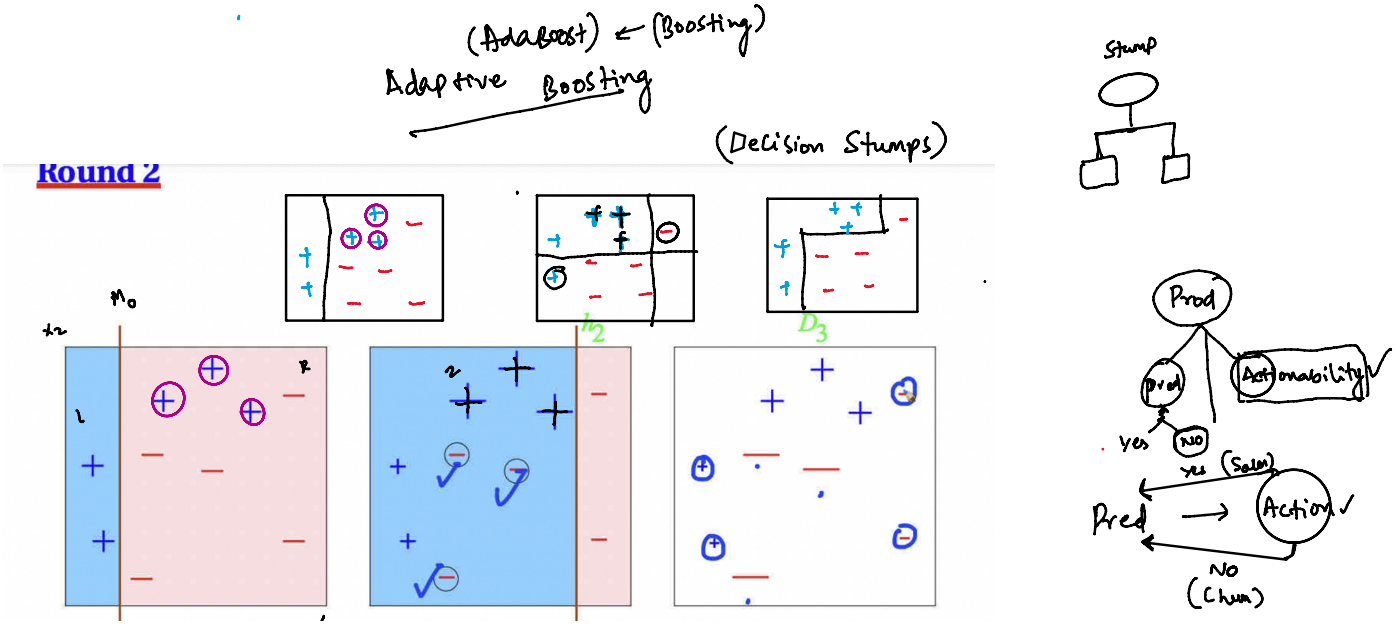
Column Sample by node

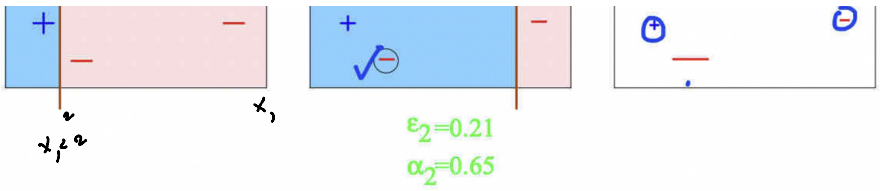


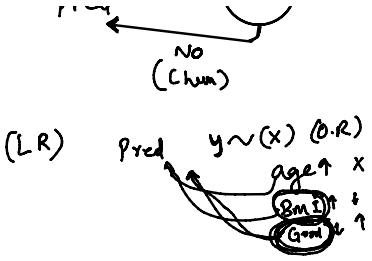
'Microsoft' '2018' Light GBM ← GOSS – (Gradient based one Sided Sampling)

M_0


M0
y M0 e0
10 9 1
20 14 6
22 17 5
M1
y M1 e
1 0.99 0.01
0
EFB → (Exclusive Feature Bundling) {f1, f2}
(Eliminate ~~son~~ features)
x1 x2 (Strongly correlated)
(x1) (~~x2~~)
f1 f2
One hot encoding
x1_A x1_B
1 0
0 1

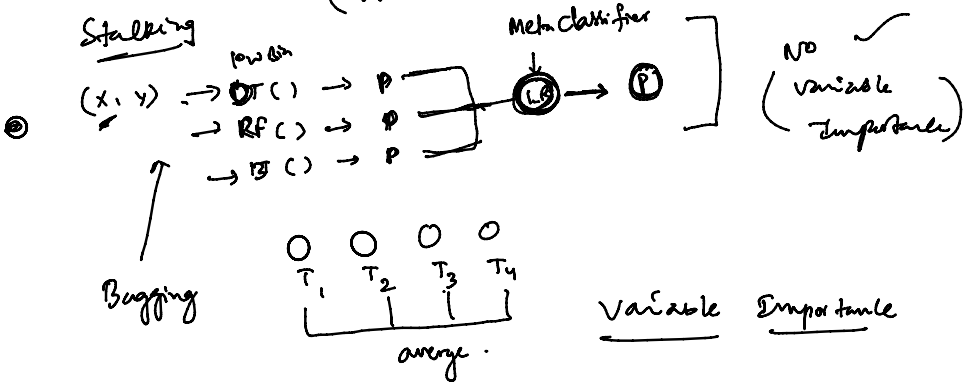

(AdaBoost) ← (Boosting)
Adaptive Boosting
(Decision Stumps)
Stump
Round 2
M0
h2
D3
Prod
Pred
Actionability
yes
NO
(Churn)
Action





Stacking ← (Kaggle)

(99%) ✓



Variable Importance

Deep learning (BlackBox) ← (Boosting, Bagging) (Lime, SHAP) ✓

(n_iter = 10)