

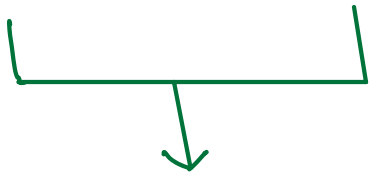
# \* Variations of GBDT

## XgBoost

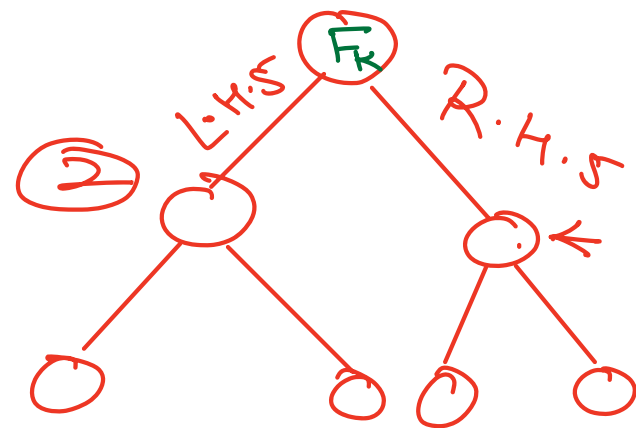
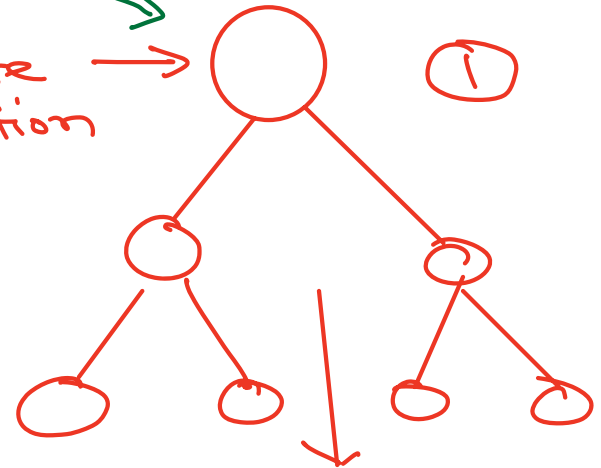
Extreme Gradient Boosting

$F_1, F_2, F_3, \dots, F_n$

Feature Selection



① Parallelization of Feature Selection



② Parallelizing

DT Learners

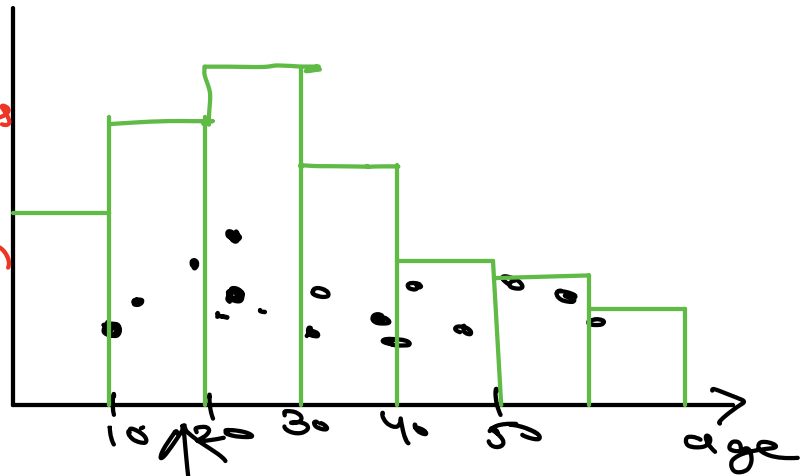
by dividing Nodes and assigning separate process to each Node

Impact: Every individual learner speeds up significantly in Training

## ③ Inbuilt binning Mechanism

### Automaticall Bin Continuous Value

- \* Less thresholds to Evaluate
- \* Noise Reduction



10  
11  
20 } 1

### Hyperparameter

- \*  $\epsilon$ TA  $\Rightarrow$  learning rate
- \* Max-depth  $\Rightarrow$  Depth of Every learner
- \* SubSample  $\Rightarrow$  Row-sampling ratio

$$S \quad \text{GBDT} + \text{Parallelization} = \text{XgBoost} \\ + \text{Binning}$$

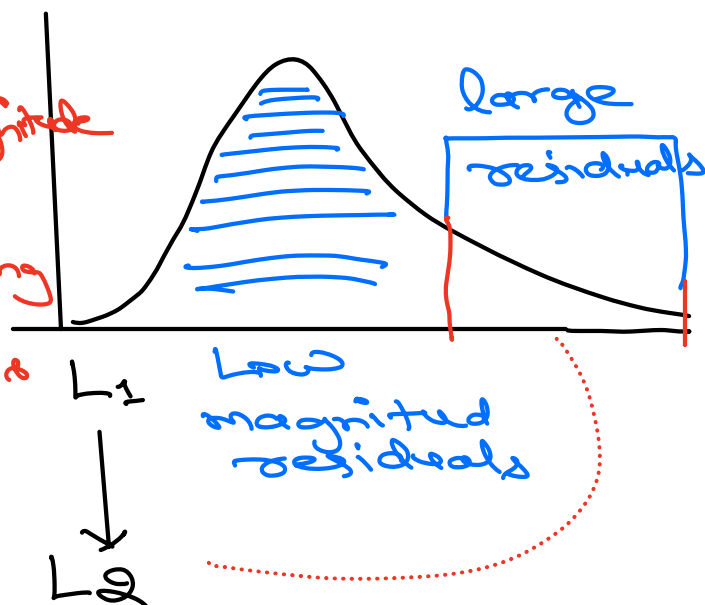
# Light GBM

\* Proposed by Researchers at Microsoft 2017

& Even more optimized than XGBoost

## 1) GOSS (Gradient based One Side Sampling)

- ① Calculate grade of all residuals
- ② Sort it acc to magnitude
- ③ Define TH threshold and sample by giving more importance to High magnitude Grads



Reduces No. of Datapoints Drastically thus speeding up overall Training

## 2) EFB (Exclusive feature)

Exclusive features Banding

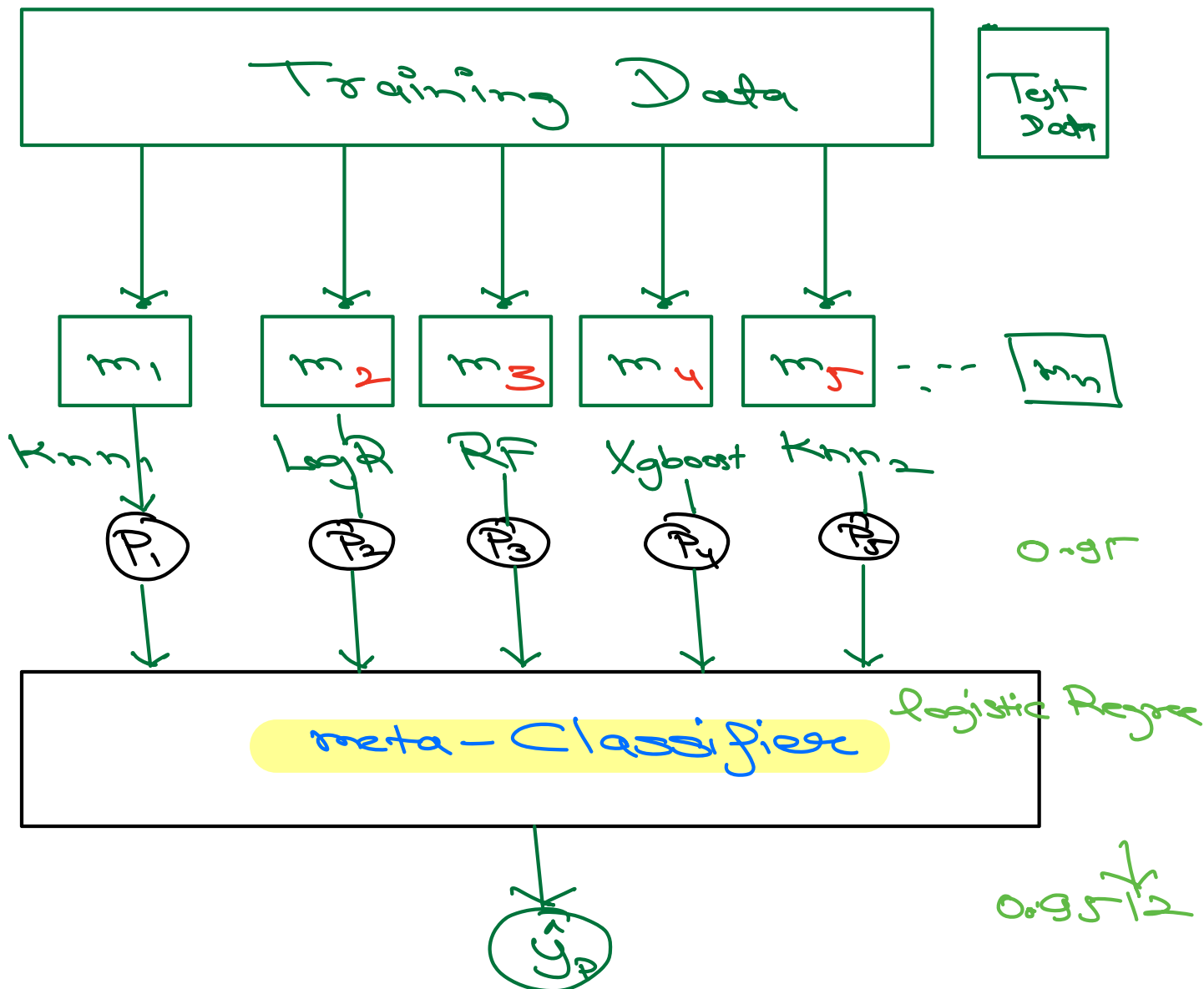
F <sub>1</sub>	m	A
	1	0
	0	1
	1	0

→

C.F
10
01
10

EFB tries to find such Exclusive  
feature pairs and group them  
↳ Dimensionality Reduction

## Stacking



\* All models are trained independently  
on full Dataset

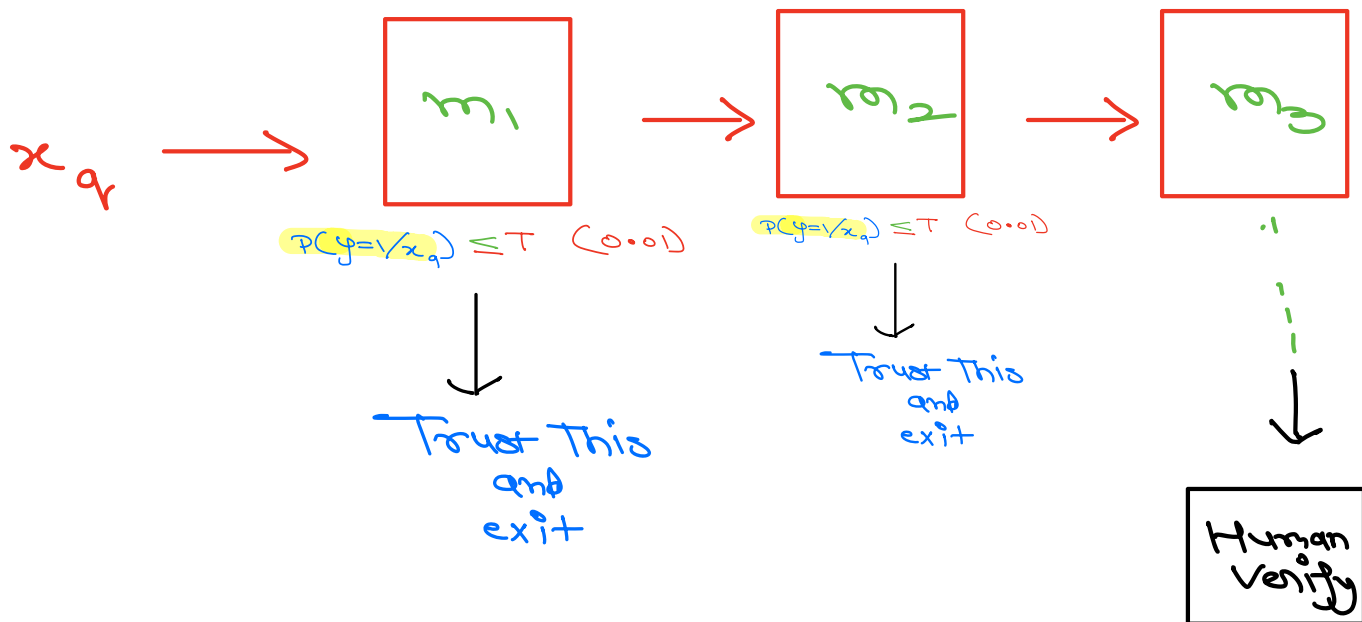
# Cascading

Credit Card Transaction Classification

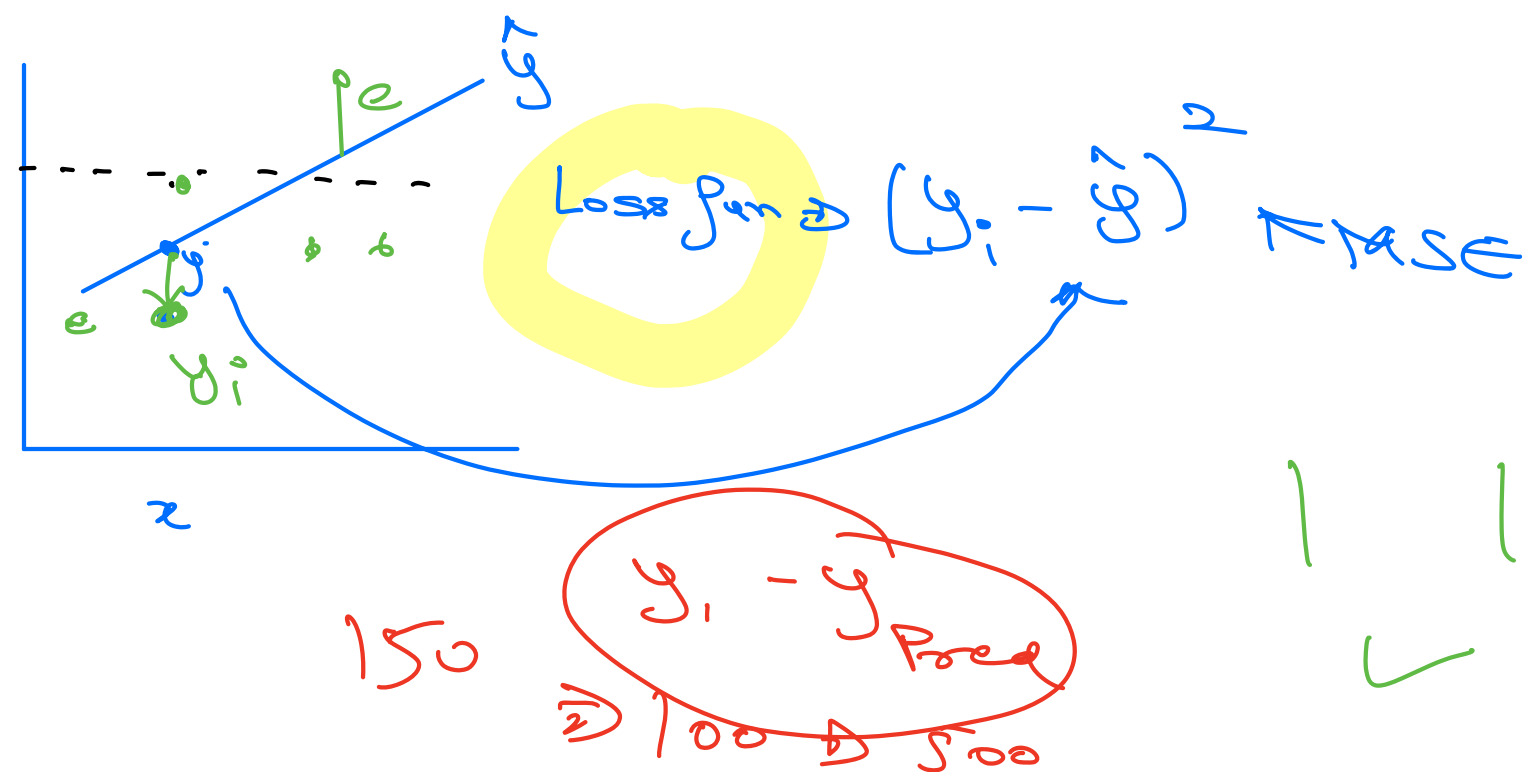
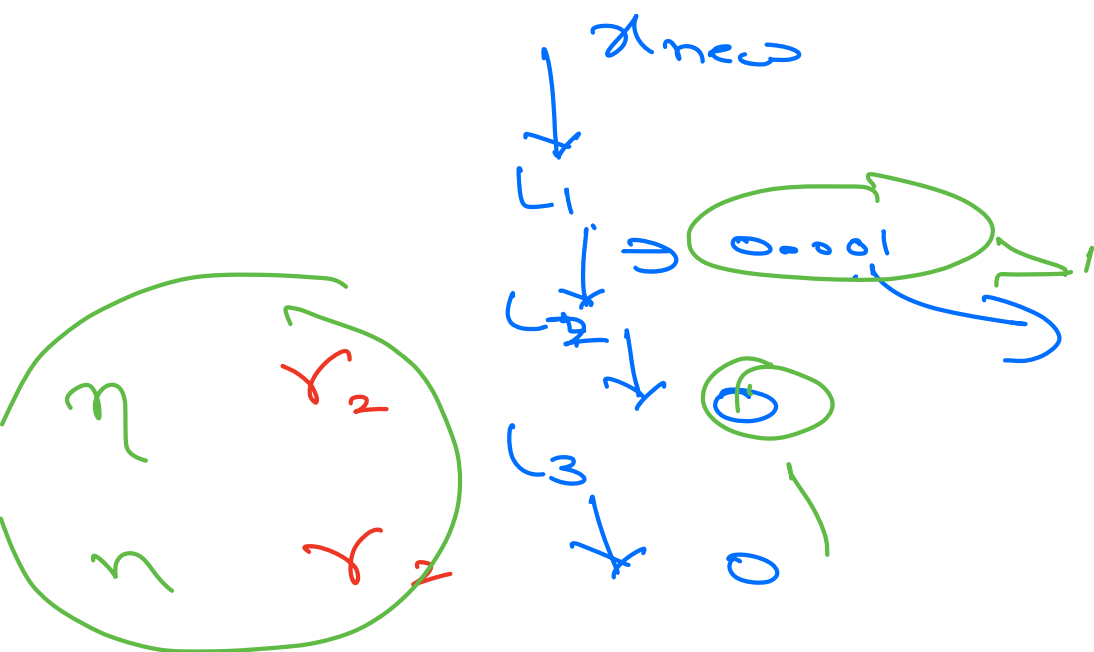
→ Fraud → Class 1

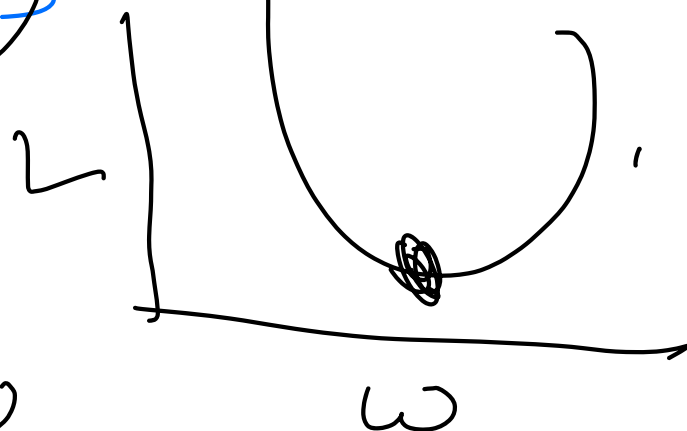
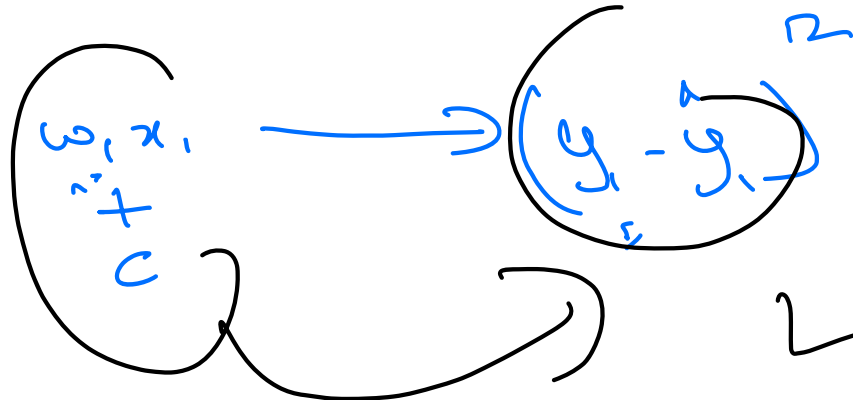
→ Valid → Class 0 (Not fraud)

Cost of Misclassification is very High



\* Cascading: Tru?





$$\frac{\partial U}{\partial y}$$

