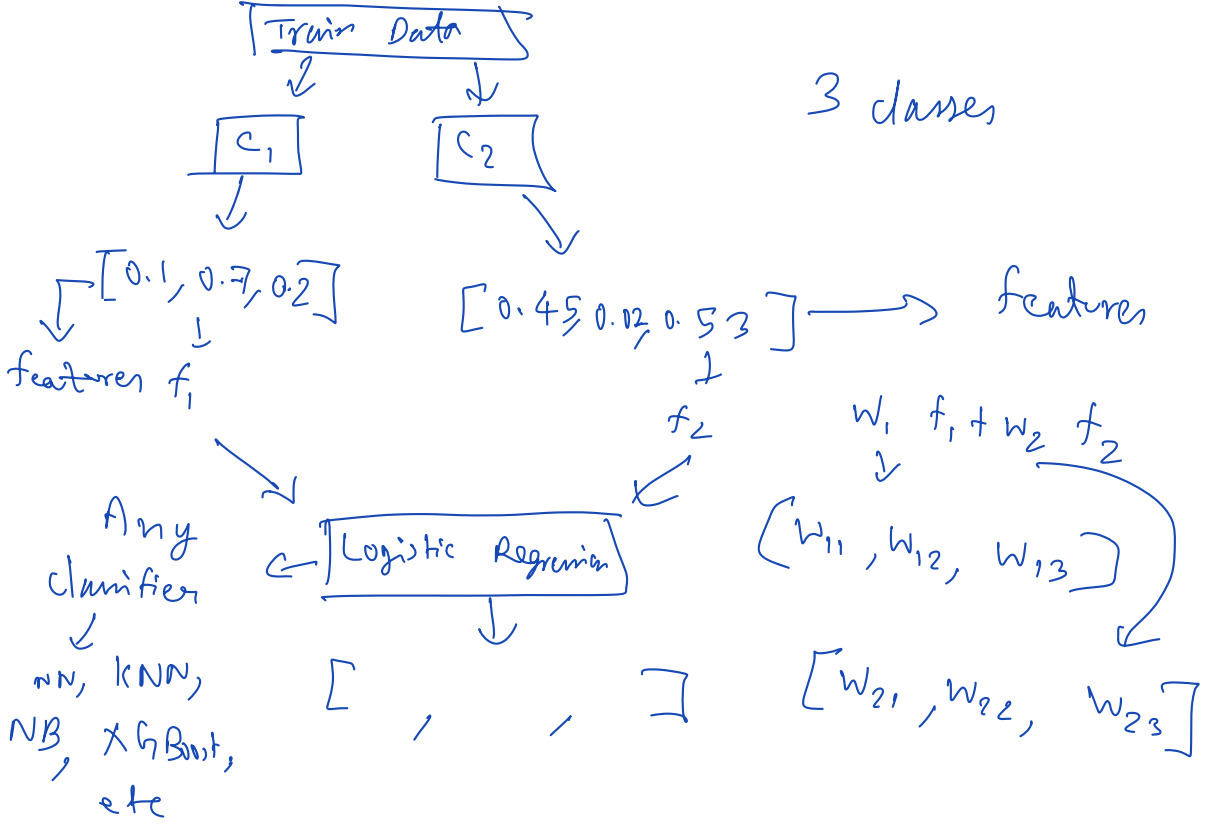
Last Class (18 Sept)

1) Quizzes
2) Recap of prob. & loss fn in Boosted Trees
3) How classification is done in Boosted Trees
4) Bias Variance Trade-off
5) How to regularize GBDT
6) Does Outlier impact GBDT
7) Use Case - EMG Signal Classification

Today's Class

1) Recap + Quizzes
2) XGBoost & Optimization
3) XGBoost Hyper-params
4) Code walkthrough of XGBoost
5) Light GBM & Optimization
6) Code walkthrough of Light GBM
7) Stacking
8) Cascading

10:27 → Resume



Logistic Regression

2 class

A → [1, 0] Target Variable

B → [0, 1]

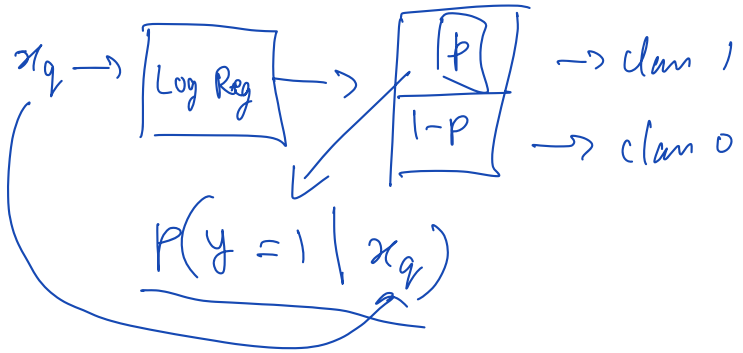
y → [1, 0]

$\hat{y}$ → [0.9, 0.1]

3 classes → A, B, C

S_1 → A : V_1 → [1, 0, 0] (1 at A, B, C)

S_2 → C : V_2 → [0, 0, 1] (A, B, C)



p → prob. of belonging to class 1

sigmoid

If $p >= 0.5$, $\hat{y} = 1$

$p < 0.5$, $\hat{y} = 0$

Threshold $= 50\% = 0.5$

If $p < 0.1\%$ $\hat{y} = 0$

Th

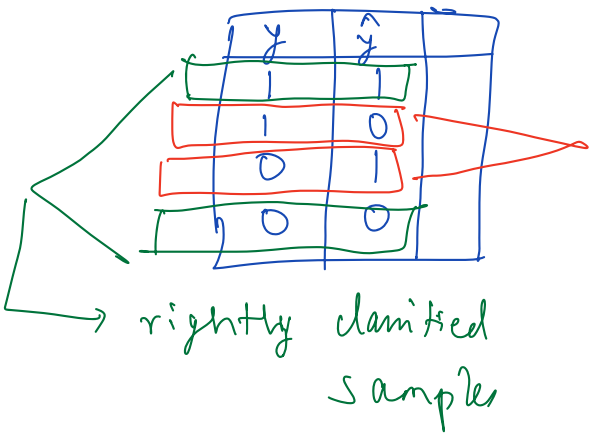
chances of fraud or prob. happening is less than 0.1%

directly classify it as non-fraud

$p >= 0.1$, $\hat{y} = 1$

Fraud → class 1 (+ve)

Non-Fraud → class 0 (-ve)



wrongly classified samples

rightly classified samples