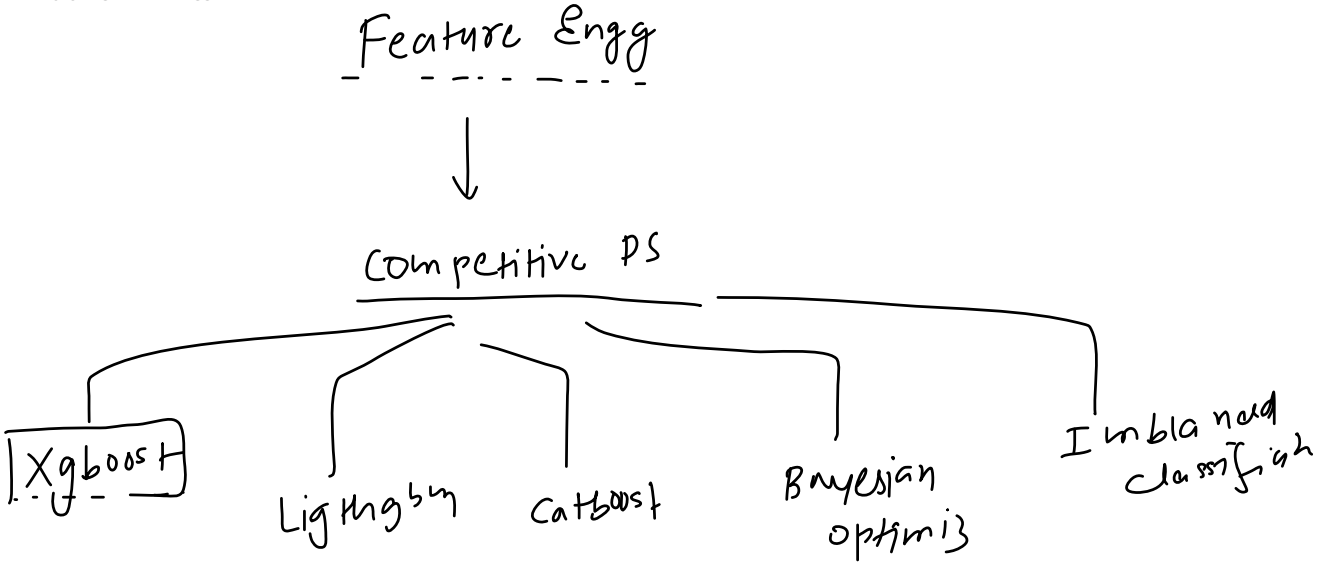
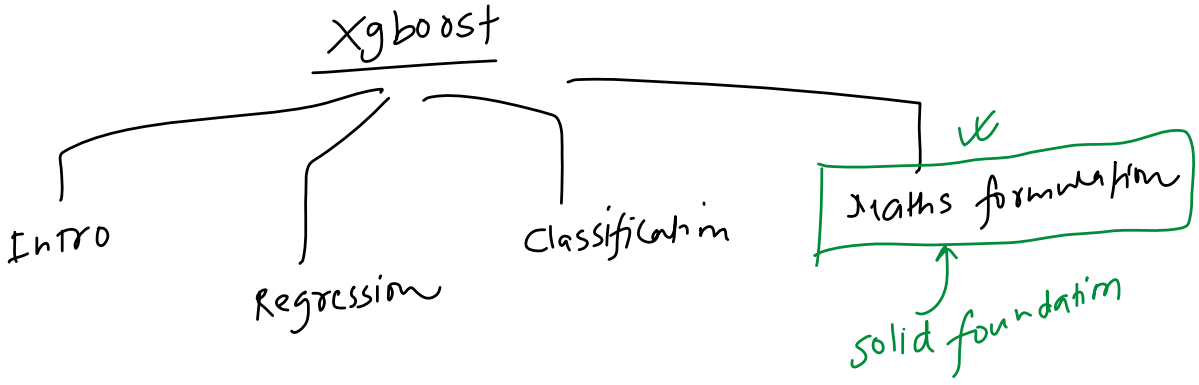
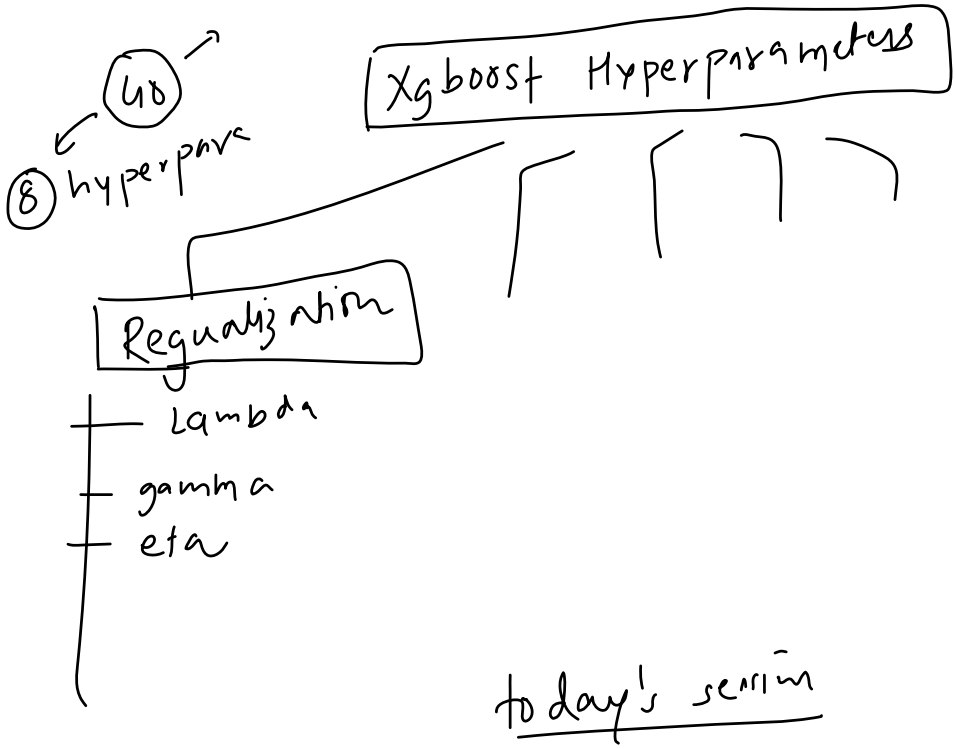
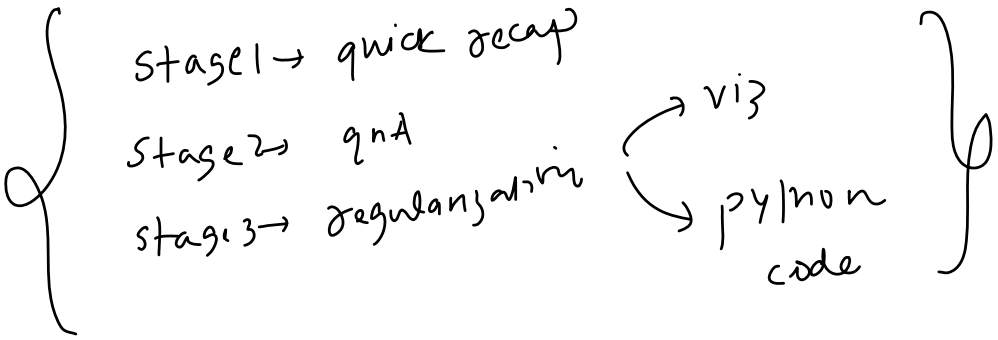
Plan of Attack

24 March 2024 12:00








stage1 → quick recap
stage2 → qnA
stage3 → regularization
viz
python code

Supervised ML
23 March 2024 15:22

cgpa	iq	package
7.2	80	3
8.1	110	4
5.6	90	12
7.3	95	19

x_i y_i

Xgboost → 2 ideas (1st idea)

model → predictions → $\hat{y}_i$ →

parametric ✓ (linear reg, lvr)

Non-parametric (DT)

mse / logloss

$L(\hat{y}_i, \hat{y}_i)$

Loss minimi

minimi

Problems with just keeping the loss function

overfitting

How do we know the predictions are any good?

(Loss function)

$L(\theta)$

Objective function

$Obj(\theta) = L(\theta) + \Omega(\theta)$

variance

Training Loss (measures how well the model fits on training data)

bias

Regularization (measures complexity of the model)

- Ridge regression: $\sum_{i=1}^{n}(y_i - w^T x_i)^2 + \lambda\|w\|^2$
  - Linear model, square loss, L2 regularization
- Lasso: $\sum_{i=1}^{n}(y_i - w^T x_i)^2 + \lambda\|w\|_1$
  - Linear model, square loss, L1 regularization
- Logistic regression: $\sum_{i=1}^{n}[y_i \ln(1 + e^{-w^T x_i}) + (1 - y_i)\ln(1 + e^{w^T x_i})] + \lambda\|w\|^2$ (reg; logloss)

Helps in Bias Variance Tradeoff

$L(\theta) + \Omega(\theta)$

Regularization
↳ parametric models → linear, neural networks

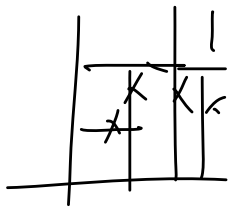
Stagewise Additive Modelling

23 March 2024 15:50

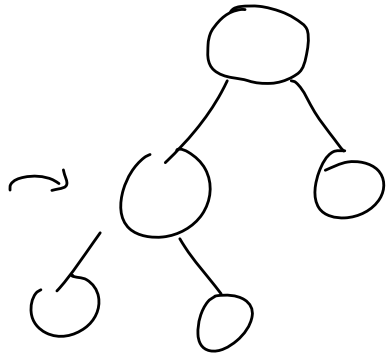
2nd interim

tree based models

Decision tree

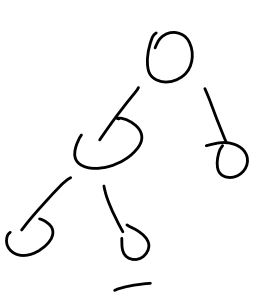


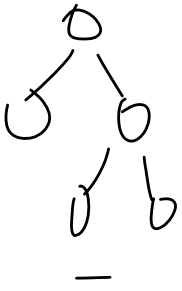
cgpa | iq | package

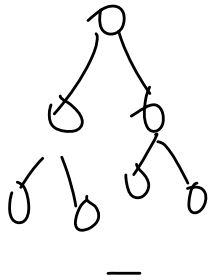


overfitting

Random forest



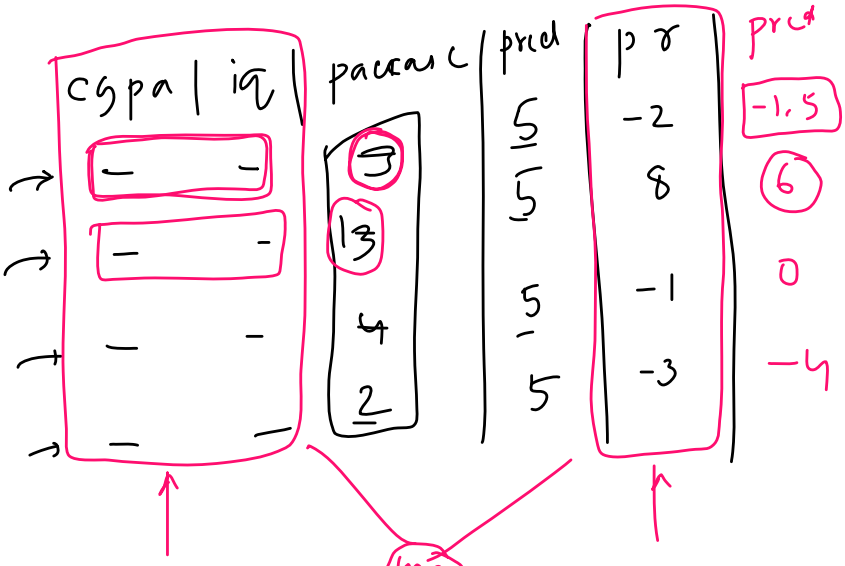


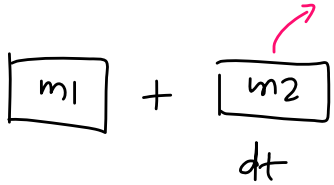


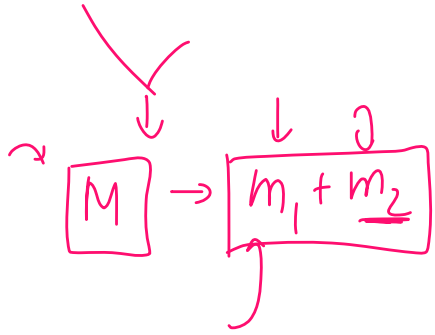
→ random forest

5 lpa

Stagewise additive modelling





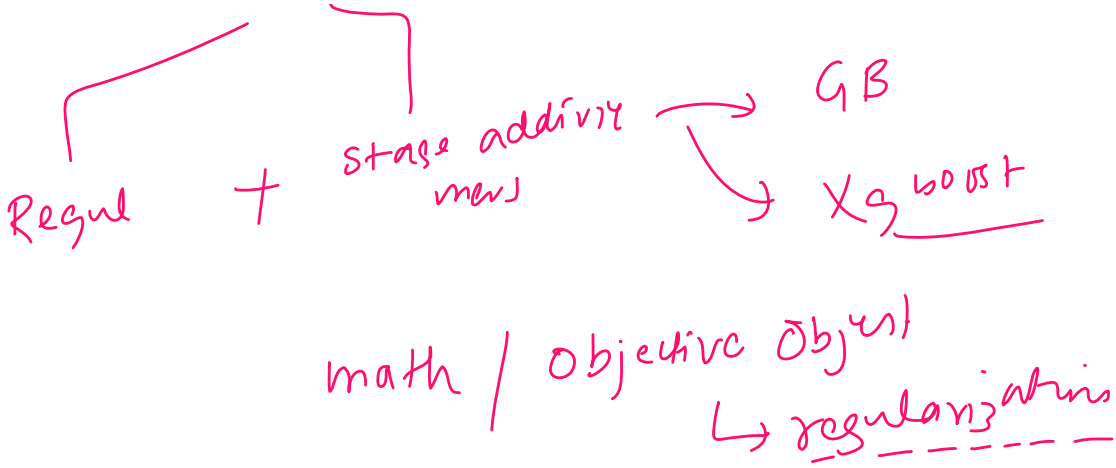


5 + (-1.5) → 3.5

5 + 6 → 11

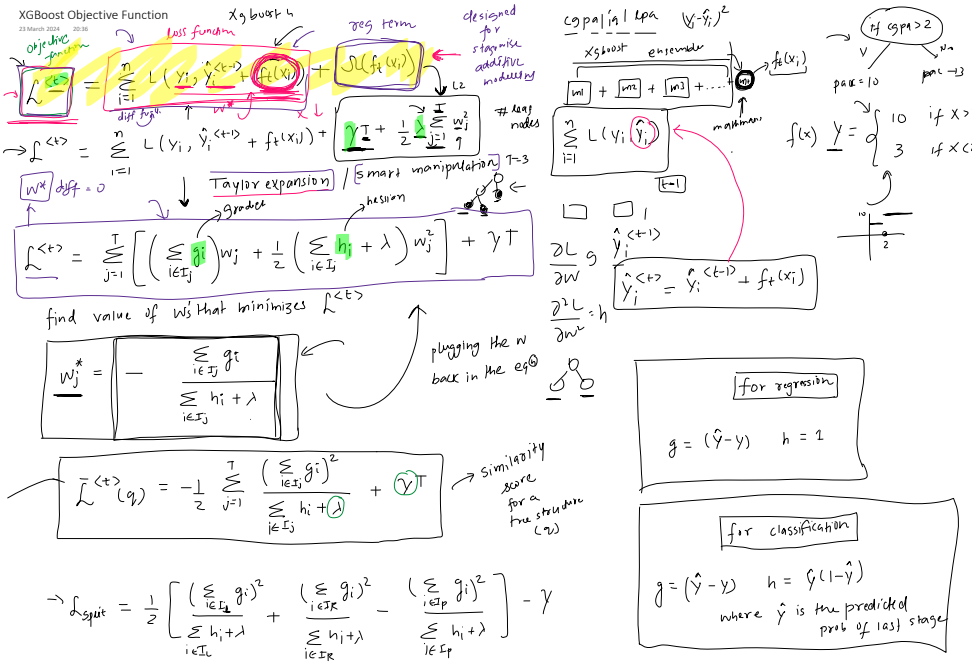
gradient boosting

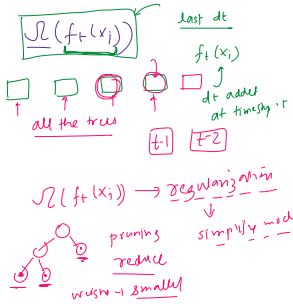
xgboost

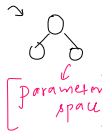

Regul + stage additive
GB
XGBoost
math / objective Objective
regularization

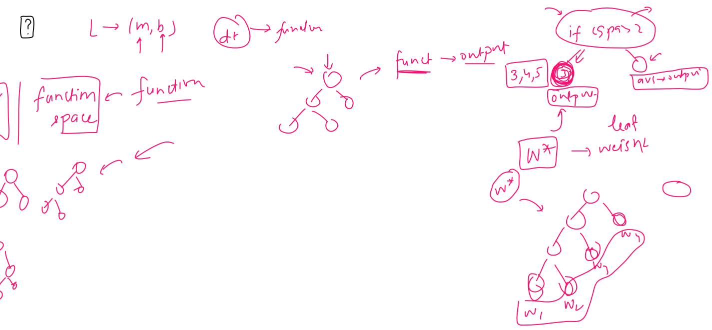
XGBoost Objective Function

23 March 2024 20:36









XGBoost High Level Overview

24 March 2024 07:18

1. Lambda (& Alpha)

24 March 2024 00:35

2. Gamma

24 March 2024 00:36

3. Shrinkage

24 March 2024 00:36

4. Max Depth

24 March 2024 00:36

5. Max Leaves

24 March 2024 00:38

6. Min Child Weight

24 March 2024 00:39

7. Subsample

24 March 2024 00:36

8. Col Subsample(ByTree/ByLevel/ByNode)

24 March 2024 00:37