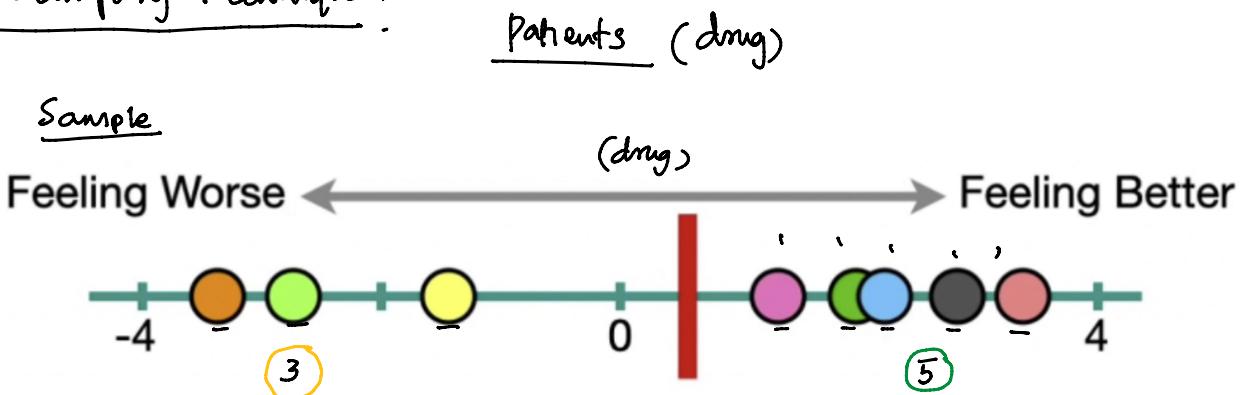


Bagging and RF

19 February 2022 20:46

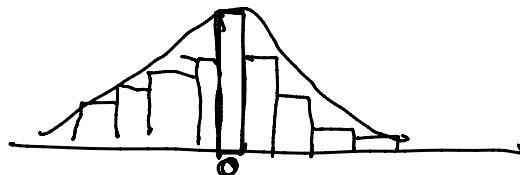
Bootstrap Aggregation

Re Sampling technique:



Experiment

Sampling distribution



Bootstrapping



drug has no effect

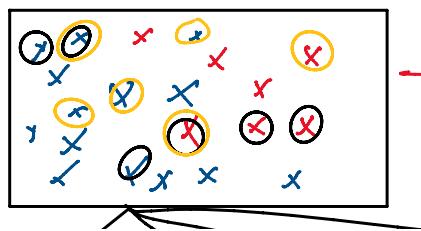
Bootstrap Aggregation: (Bagging)

Re sample techniques

With Replacement ✓

Without Replacement

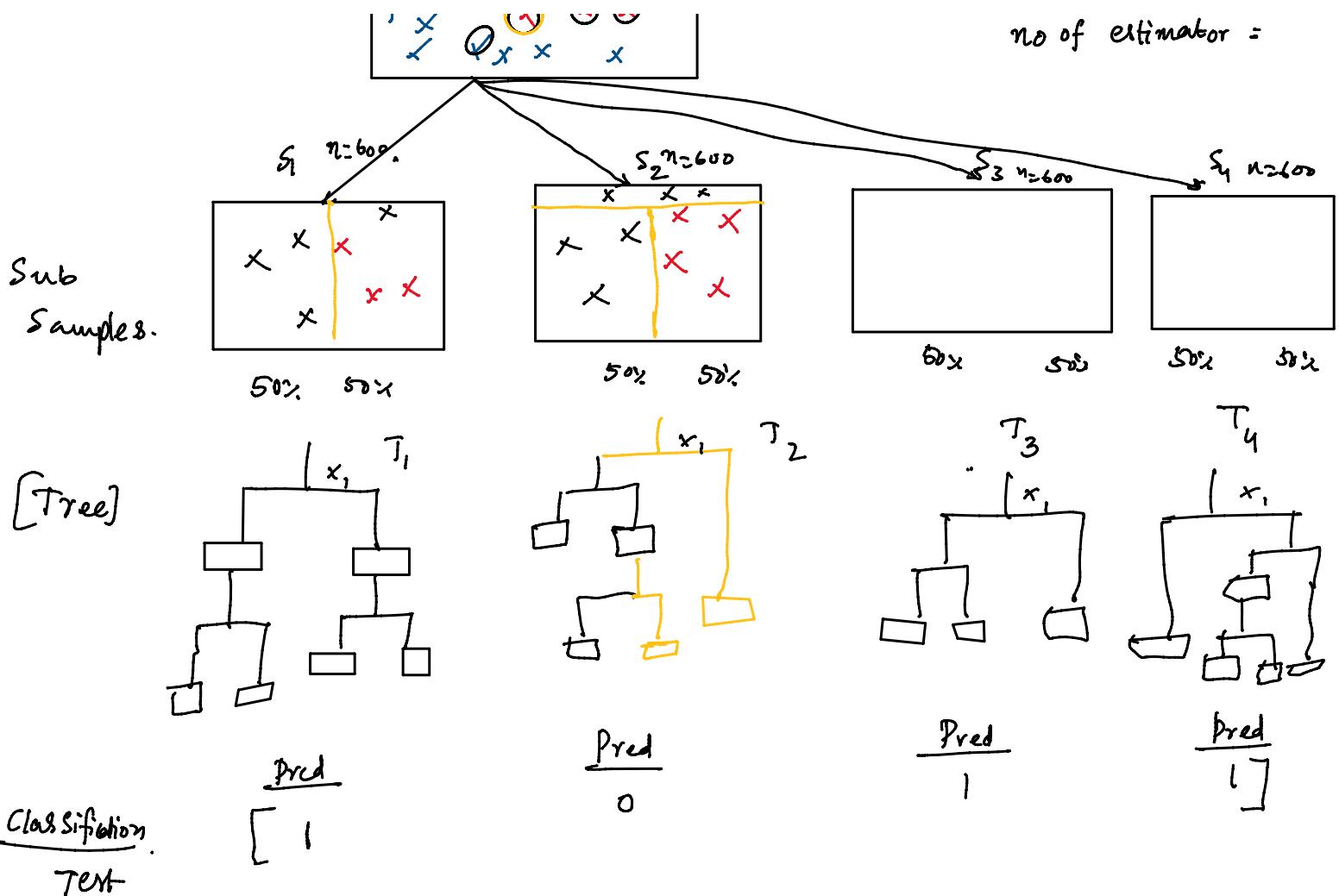
Sample.



0 - 80% $n = 1000$

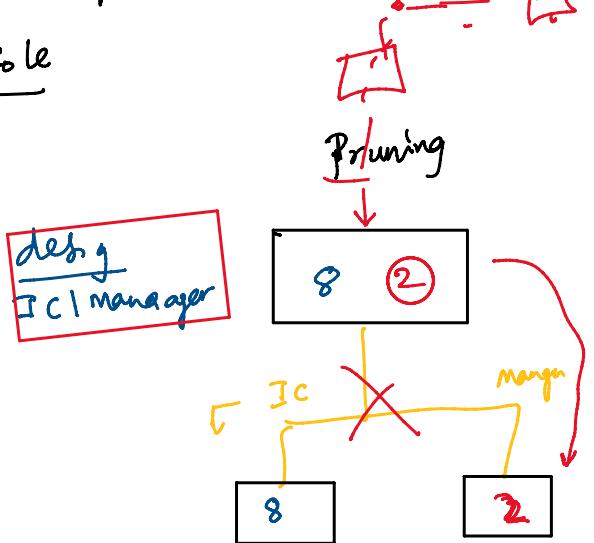
1 - 20%

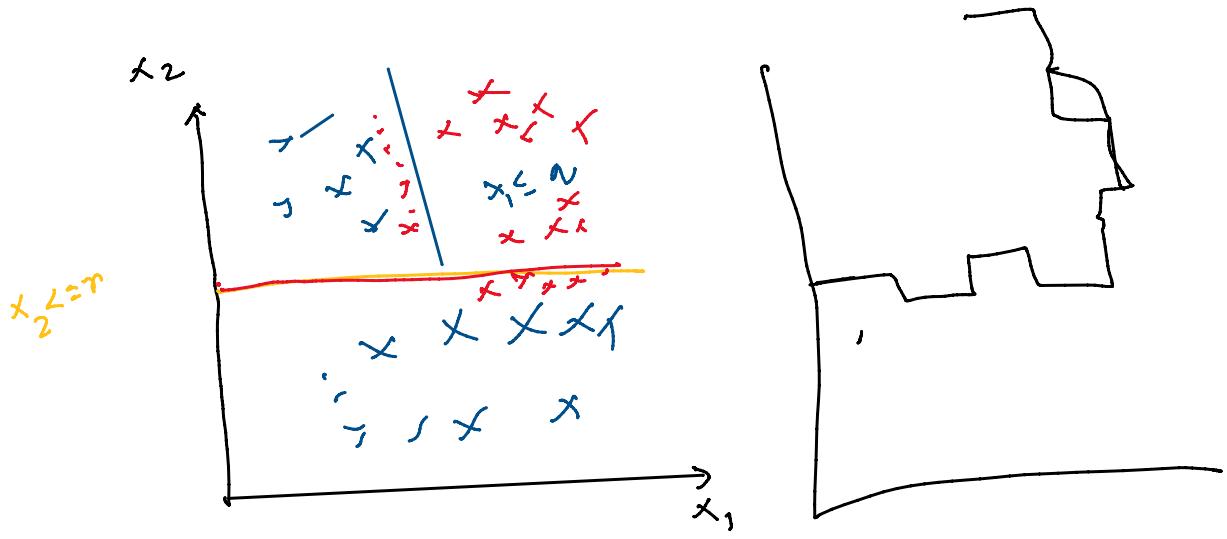
no of estimator =



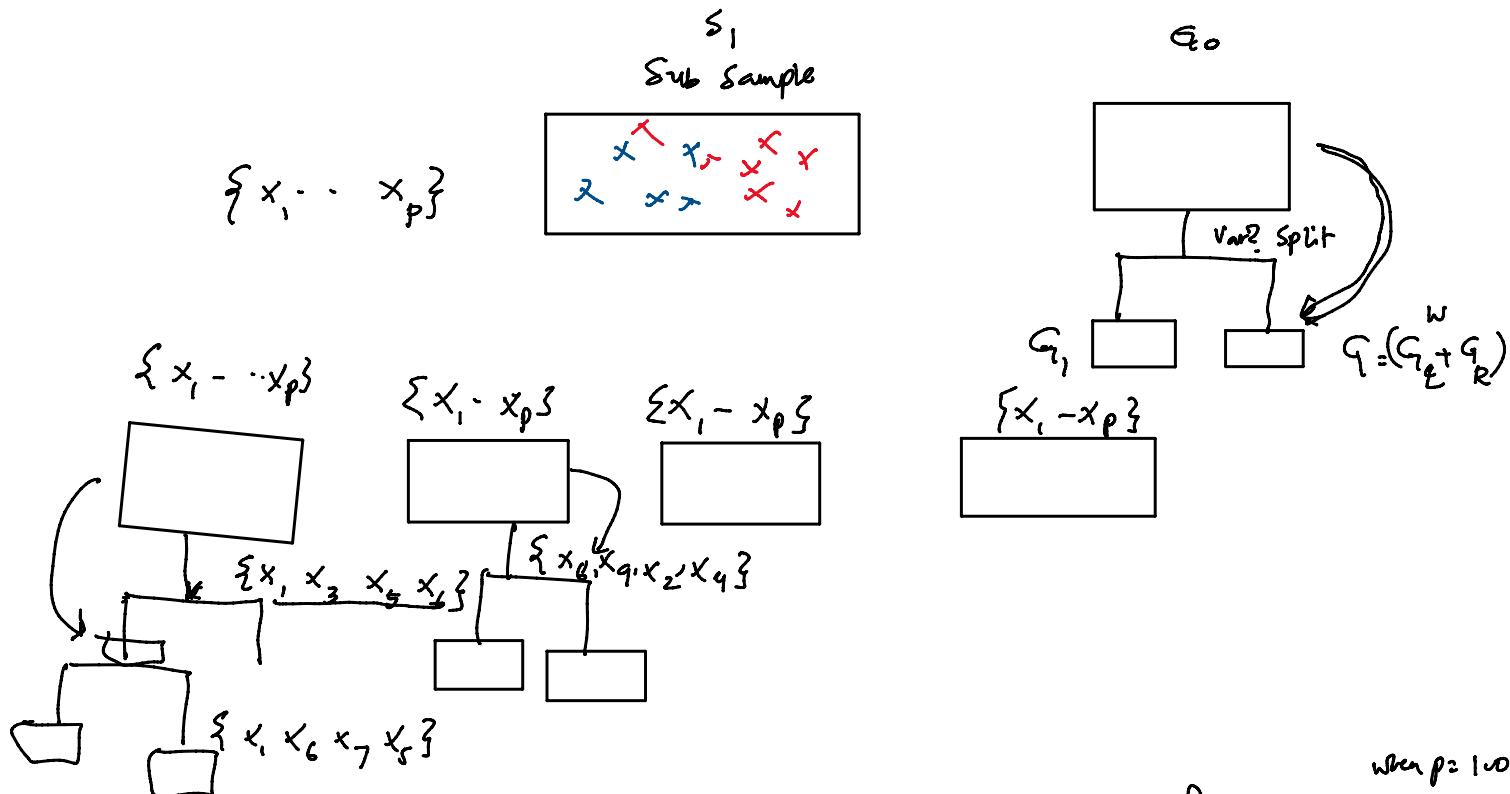
aggregation - Classification - Voting - majority
 Regression - Averaging - mean

Ensemble





Random forest (Bagging + Variable Sampling)

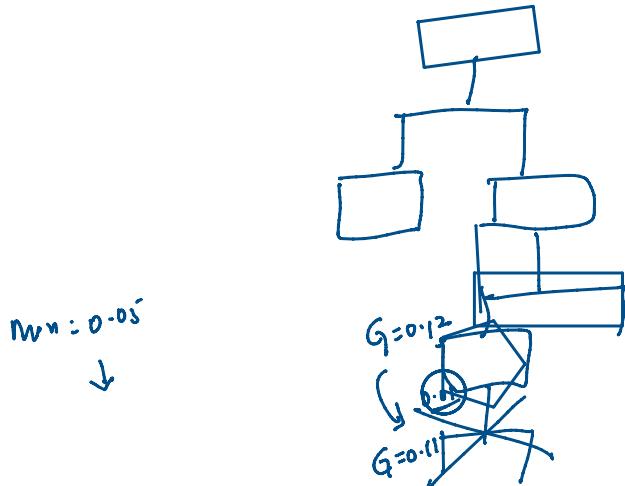
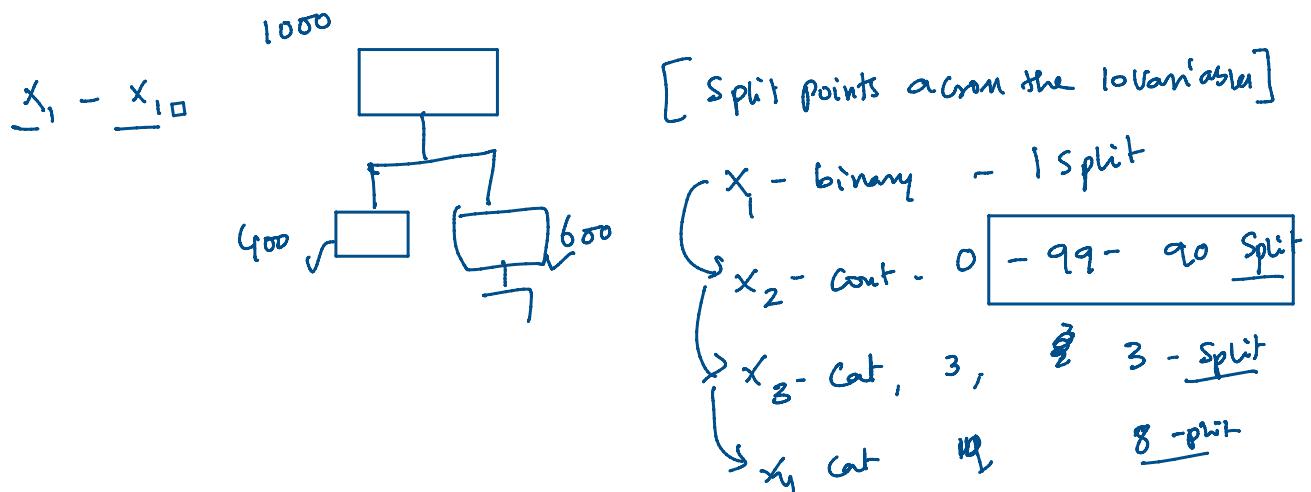
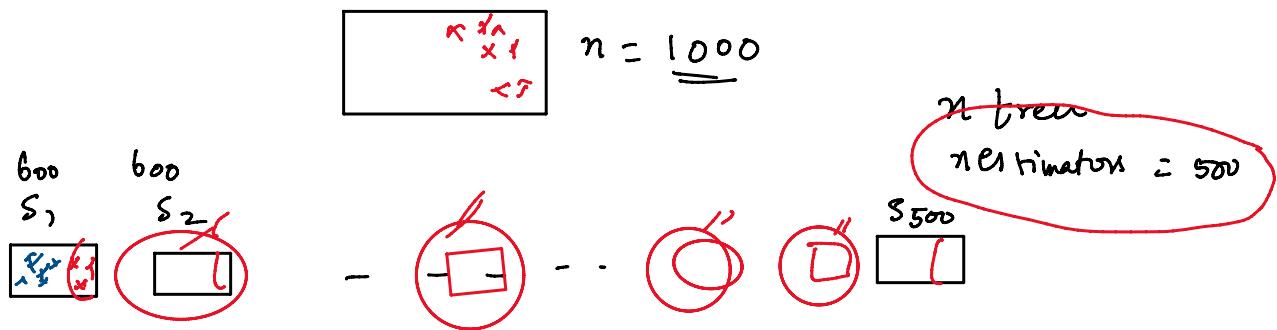


Regression $\hat{P} = \frac{m}{3} = \frac{100}{3} = 33\leftarrow$ (max_features) $m=3y$ when $p=100$

Classification $= \sqrt{P} = \sqrt{100} = 10\leftarrow$ Classification.

$m=10, P=100$

$$n = 50000$$



Hyperparameter tuning

Parameters

✓ Optimized
Hyperparameters $\theta_{optimal}$

parameter
 ↴
 model centric
 ↴
 linear reg. $(\beta_0) + (\beta_1)$

✓ Optimized
Hyperparameter (User Centric)
 Regularization $\lambda = 1$
 Learning $\alpha = 0.01 \approx 0.05$ (user)

→ n estimators = [10, 25, 50, 100, 250] ? 5.
 → max-depth = [5, 6, 7, 8, 10, 12] 5
 → max-features = [6, 7, 8] 3
 → criterion - 'gini' 'entropy' 2
 ↴
 ↴
 ↴

180

Optimization (Hyperparameter)

- Grid Search CV — Grid of Hyperparameters
- Random Search CV

- ① n_estimators=10, max_depth=5, max_features=6, criterion=gini ✓
- ② n_estimators=25, max_depth=5, max_features=6, criterion=gini ✓
- ✓
- ✓
- ✗

150

Scoring metric = Accuracy ↪ best ✓

Random Search ✓

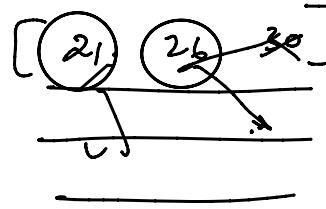
faster, most

para

n_estimators

max_features

Max_depth



start dist
norm (r vs)