

Bagging and RandomForest

10 December 2022 20:55

Ensembles

$\{ \text{Base Models} \}$ + $\{ \text{Combine them} \} \rightarrow (\text{Better than a Single model})$
one single

Text Book [1 - 20] → Learn → Test → Score

Group of friends [1 - 4] [5 - 8] [9 - 12] [13 - 16] [17 - 20] - Learning.
Base: 1 2 3 4 5

Test (combine learning) → {① ② ③ ④ ⑤} → Score ←

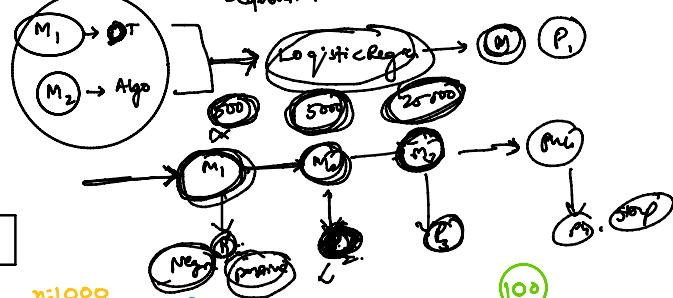
Ensembling Techniques

- Bagging ✓
- Boosting →
- Stacking →
- Cascading →

'Bootstrapped aggregation'

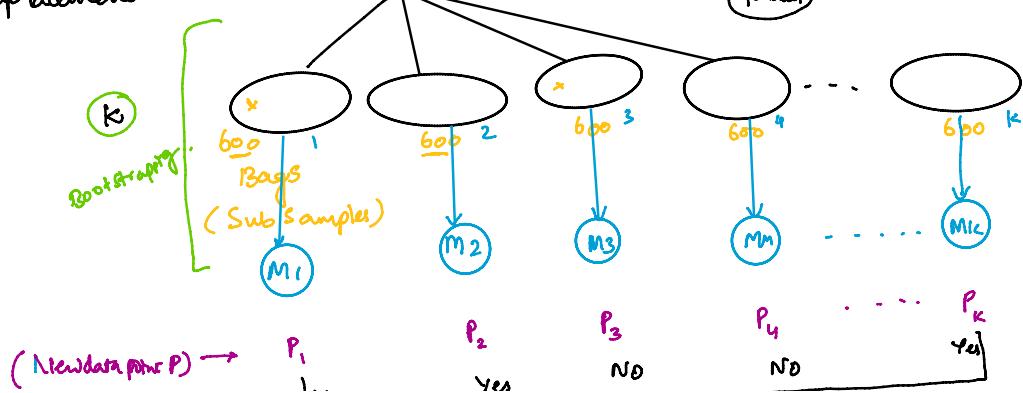


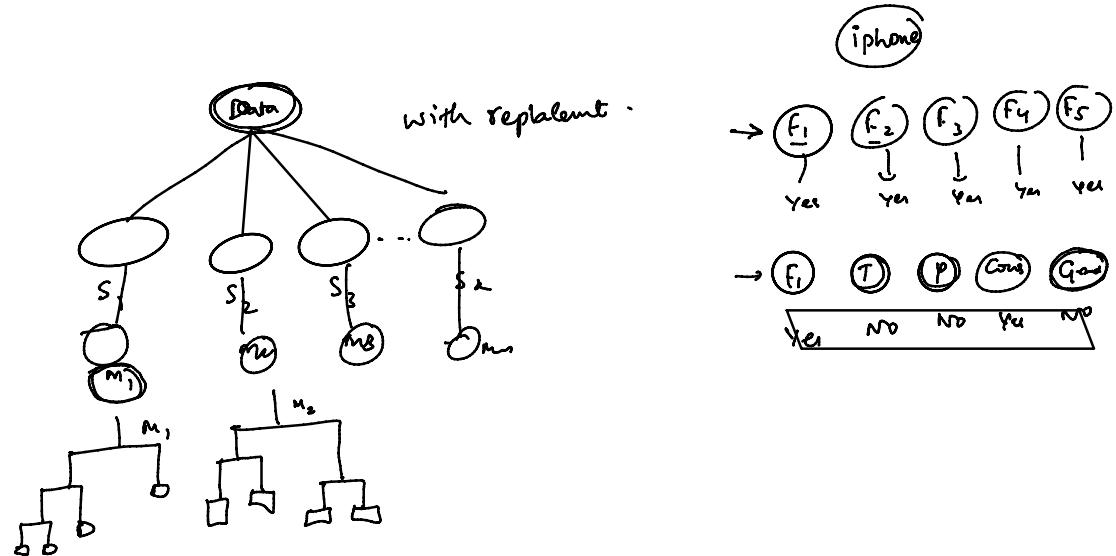
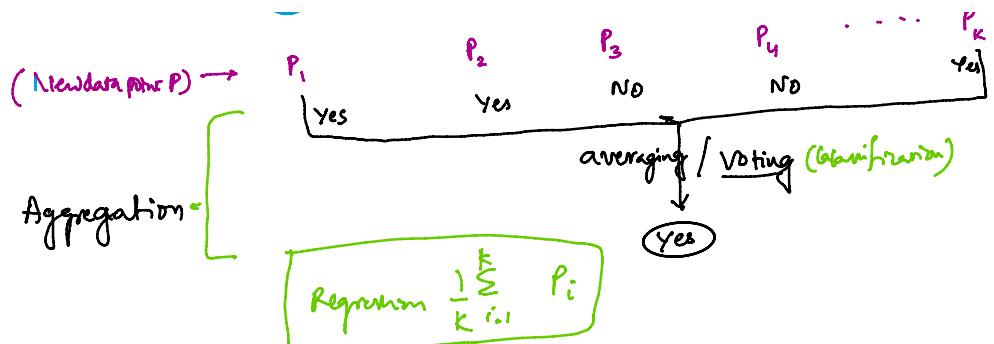
Sequential



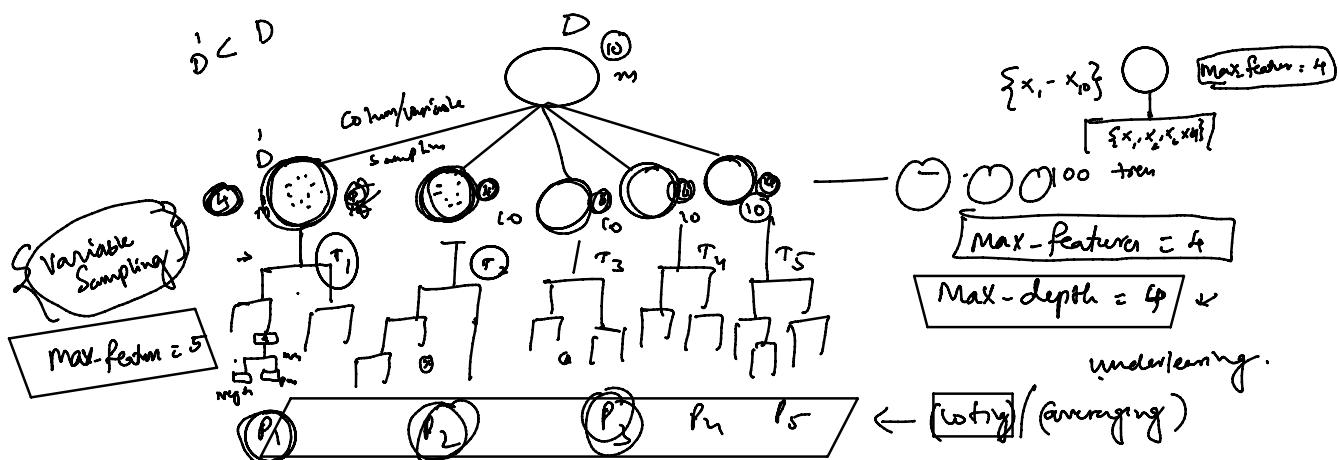
Bagging.

Resampling method.
→ With replacement



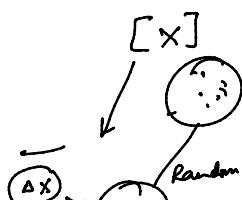


Bagging + Variable Sampling (Random Forest)

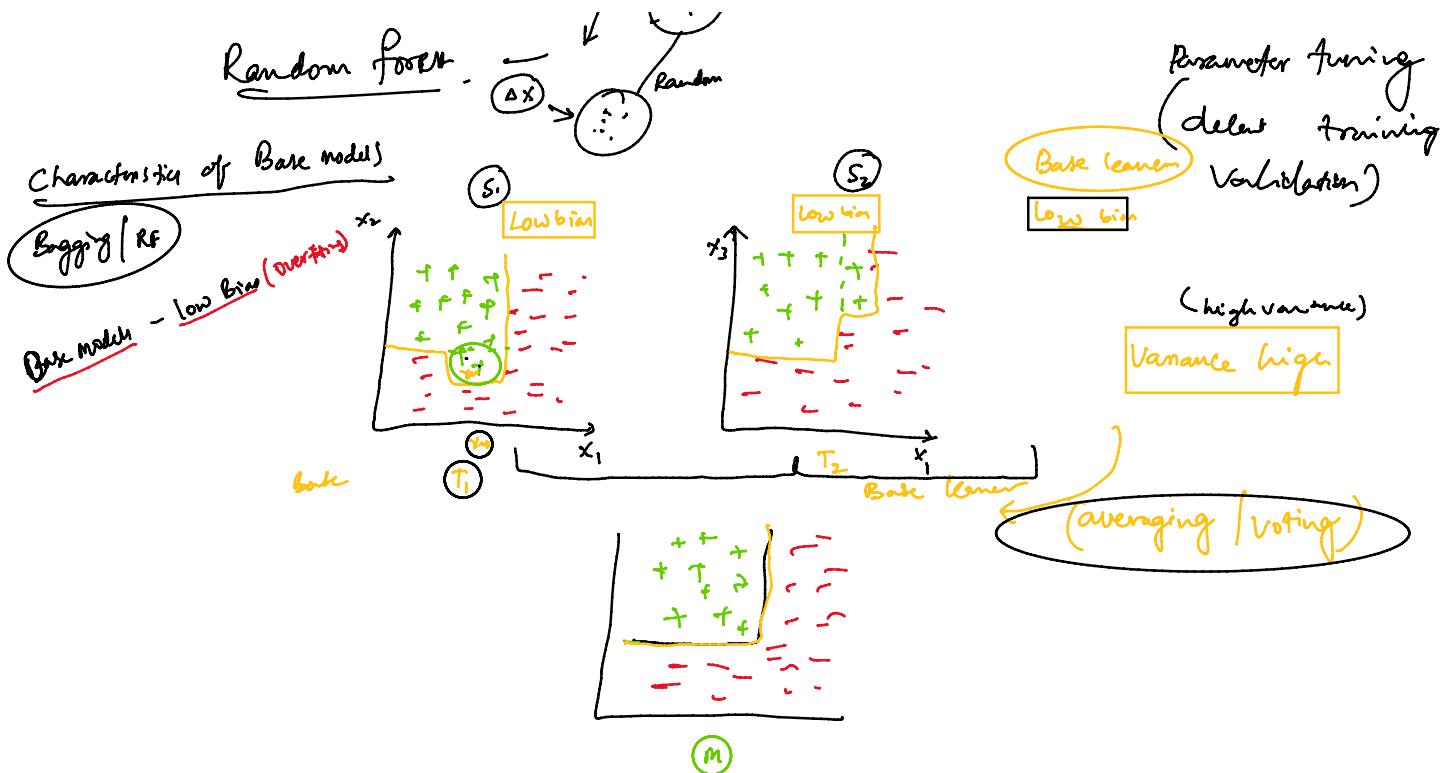


Bagging (low variance)

Random forest



Parameter tuning



RF

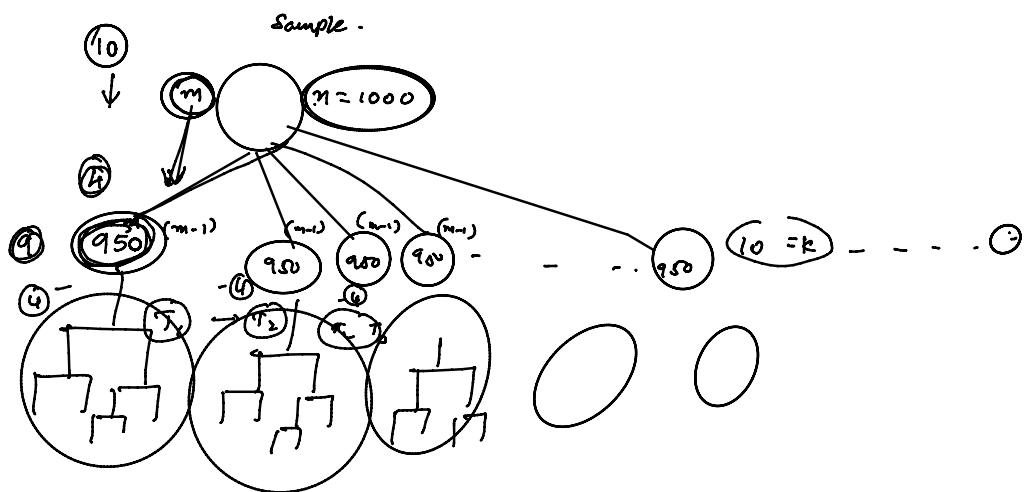
(Bagging)

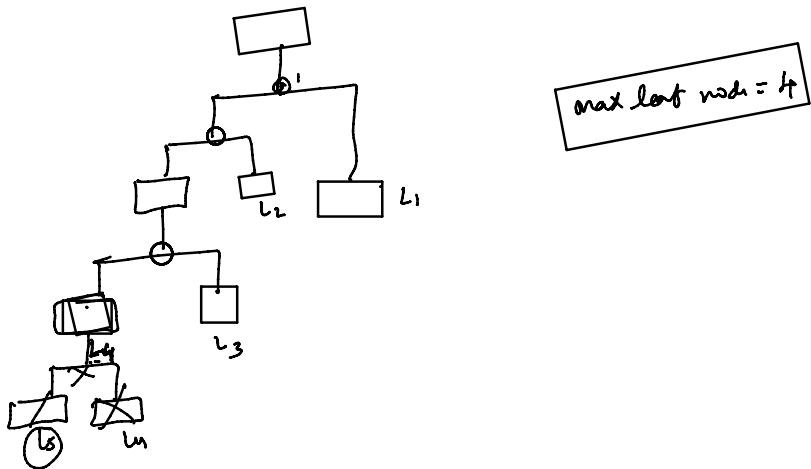
Row Sampling + Variable Sampling.

Basic learners (low bias, high variance)

Aggregation (Variance \downarrow , mean shift (bias))

\downarrow
model (Bias and Variance) ✓





C.CP → Complexity Parameter

pruning

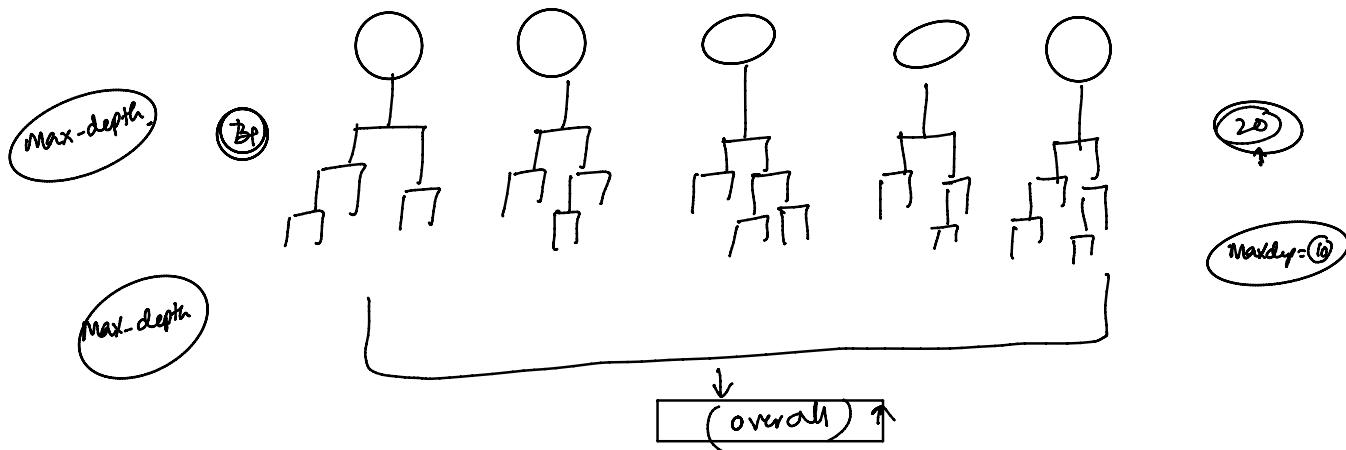
CCP- α ⇒

Tree large (\downarrow depth) complexity high.

Cost \rightarrow (loss + $\alpha \cdot$ depth of tree) Tree form $\leftarrow \alpha = 1$

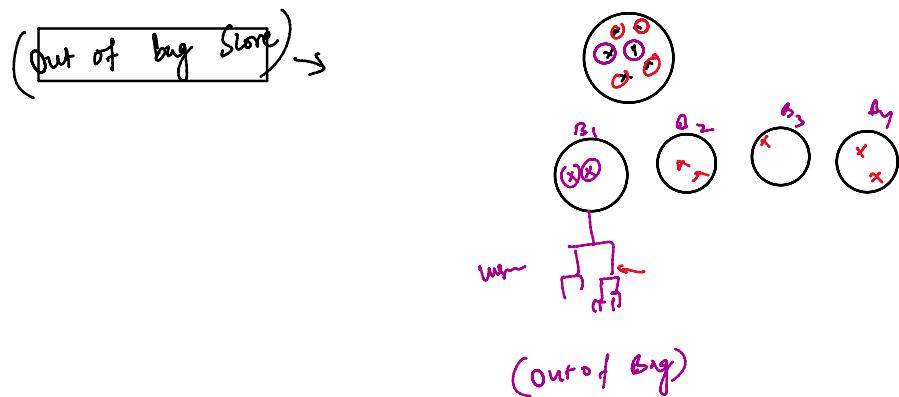
$$36 + 1 \cdot 4 \leftarrow \begin{matrix} 24 \\ 40 \end{matrix} \leftarrow \text{dep} = 1$$

$(25) + (40) \leftarrow (41) \leftarrow$ Tree



$n=10^0$

loss \rightarrow weight value

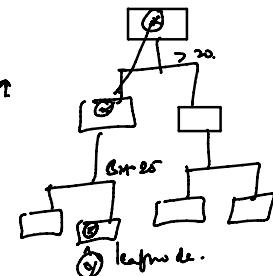


Model interpretability

Logistic Regression

Decision Tree

$\hat{y} = \frac{1}{1 + e^{-(w_0 + w_1 x_1 + w_2 x_2 + \dots)}}$

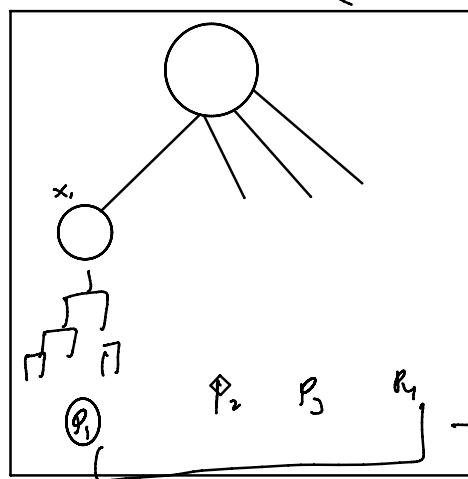


(Bagging & rf)

$\{x_1, x_2, x_3, x_4\}$ \hat{y} ?

α

(Black Box)



P

(Mean decrease impurity)

Feature importance

x_1

x_2

x_3 —————.

A diagram showing a circle with a horizontal axis passing through its center. A point labeled x_M is marked on the left side of the circle, located below the horizontal axis.

(Var import)

