Random Forest

Sunday, December 25, 2022 8:06 PM

Random forest

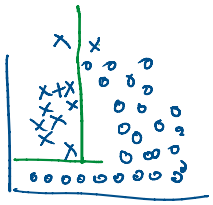
Coding
- # Scaling ✓
- # Dummies ✓
- # outlier ✓
- # Missing ✓
- # confusion Matrix

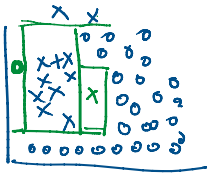
Dataset ≡ Linear, Log, Random, DT

Model
- overfit -
- underfit -
- bestfit -

• Disadvantage of Decision Tree

• overfit





Mugging up Data

30 Machi
- 29 Deep
- 1 Deep

1 Deep

• unstable

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Tennis
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

- Mathematical : Entropy, IG, → Expensive
- Greedy algorithm
  - current — It want the best

    : Mobile

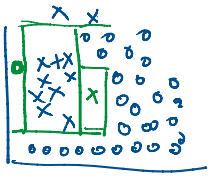
    ↳ Drugs } Best → Long Run Bad

Solution → Random forest

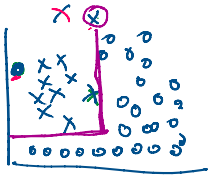
↓

{Overfit

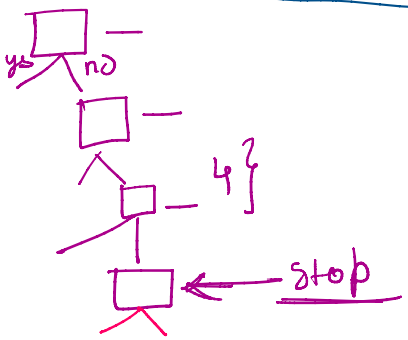
- Truncation : Stop the tree while growing
- Pruning : Cut the tree after it has grown



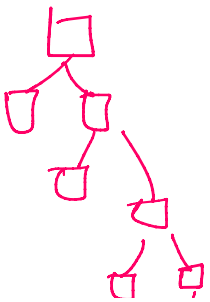
Stopping condition

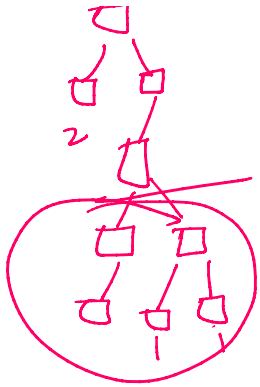


- Max depth
- Min Points } 20
- Pure leaf



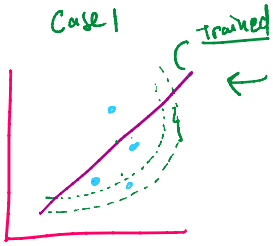
Pruning





Bias

Training error



✓ Bias : High

Variance : Low

Bias : Low

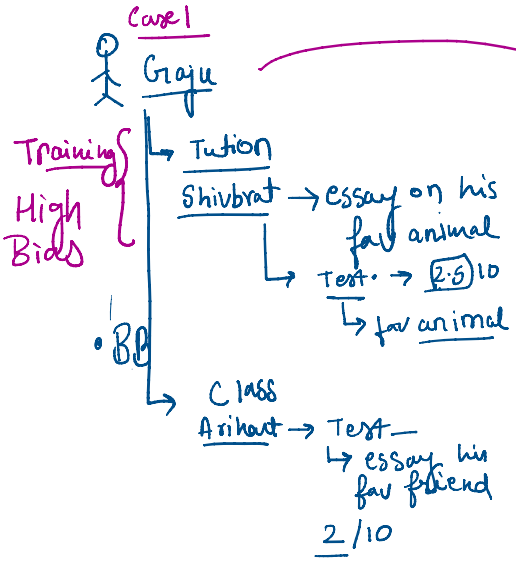
Variance : low

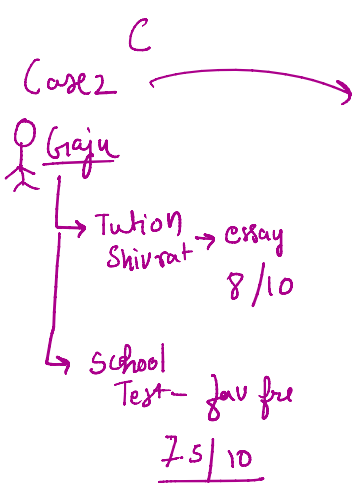
Variance

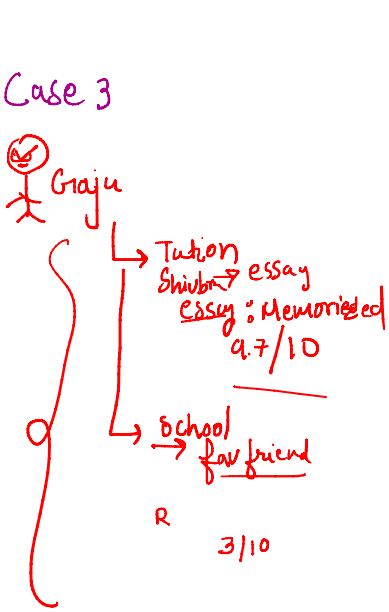
Change in Model if your data is changed

Bias : Low

Variance : High







Chatur

Trainig — 24%.

Test — 20%.

Trainig → 85%.

Test → 82%.

Training — 95%.

Training - 24%
Test - 20%
Underfit

Training → 82%
Test → 82%
Best fit
MB

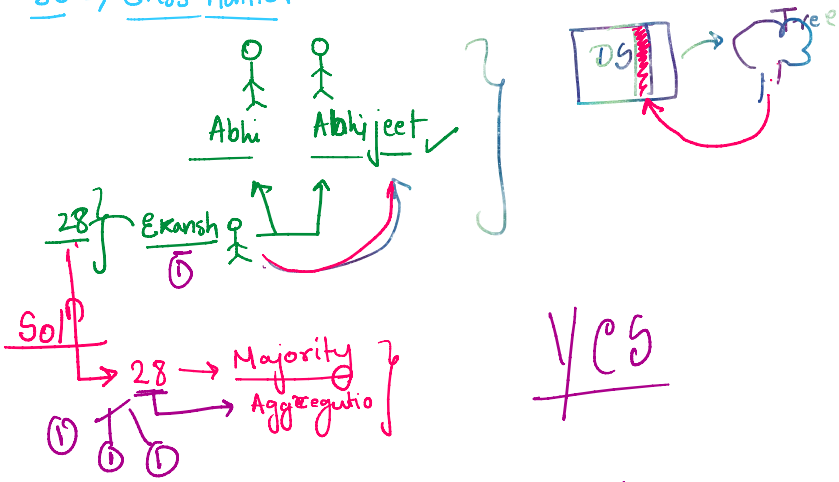
Training - 95%
Test — 72%
overfit.
JB

Decision Trees → Gets overfit

Bias → low
Variance - High

Random forest

30 → Class Monitor



VCS

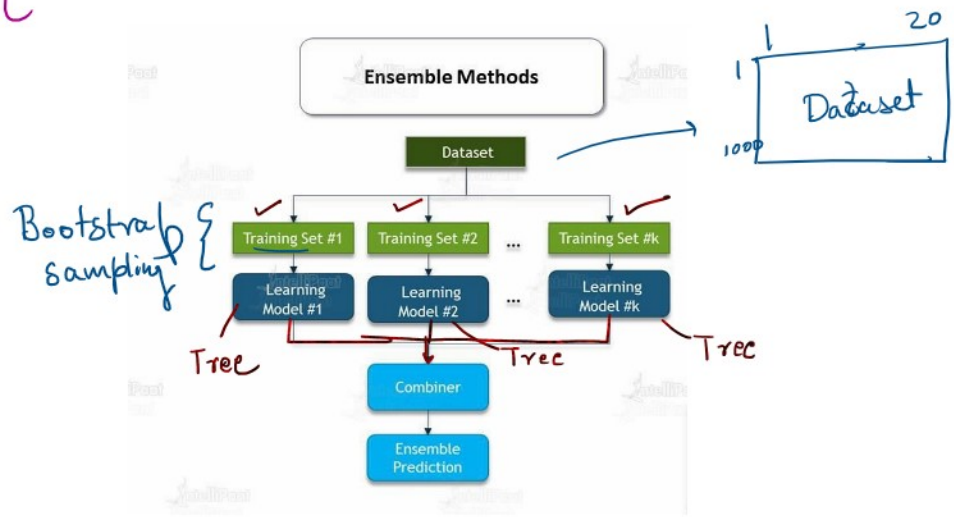
Ensemble Techqunes

→ Boosting

→ Bagging

g

↳ Bootstrap Aggregating

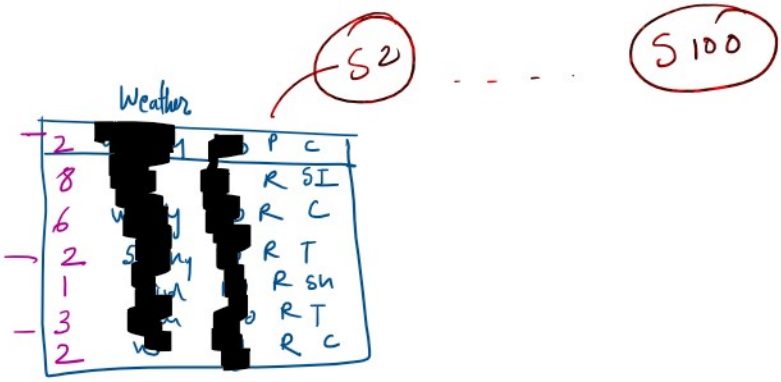


Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

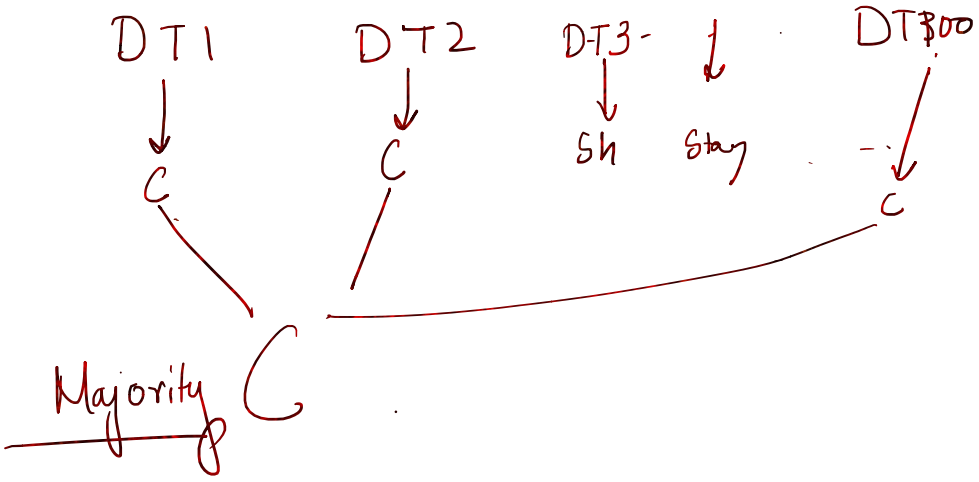
7
5
9
10
8
10
3

Randomly select the Columns without Replacment

Rows { Sampling with Replacment } Bootstrap Sampling



DT1 DT2 DT3 ... DT100



Random forest

↓

Bootstrap Tech Random

↓

Collection of Tree

Bagging helps in reduction of Variance