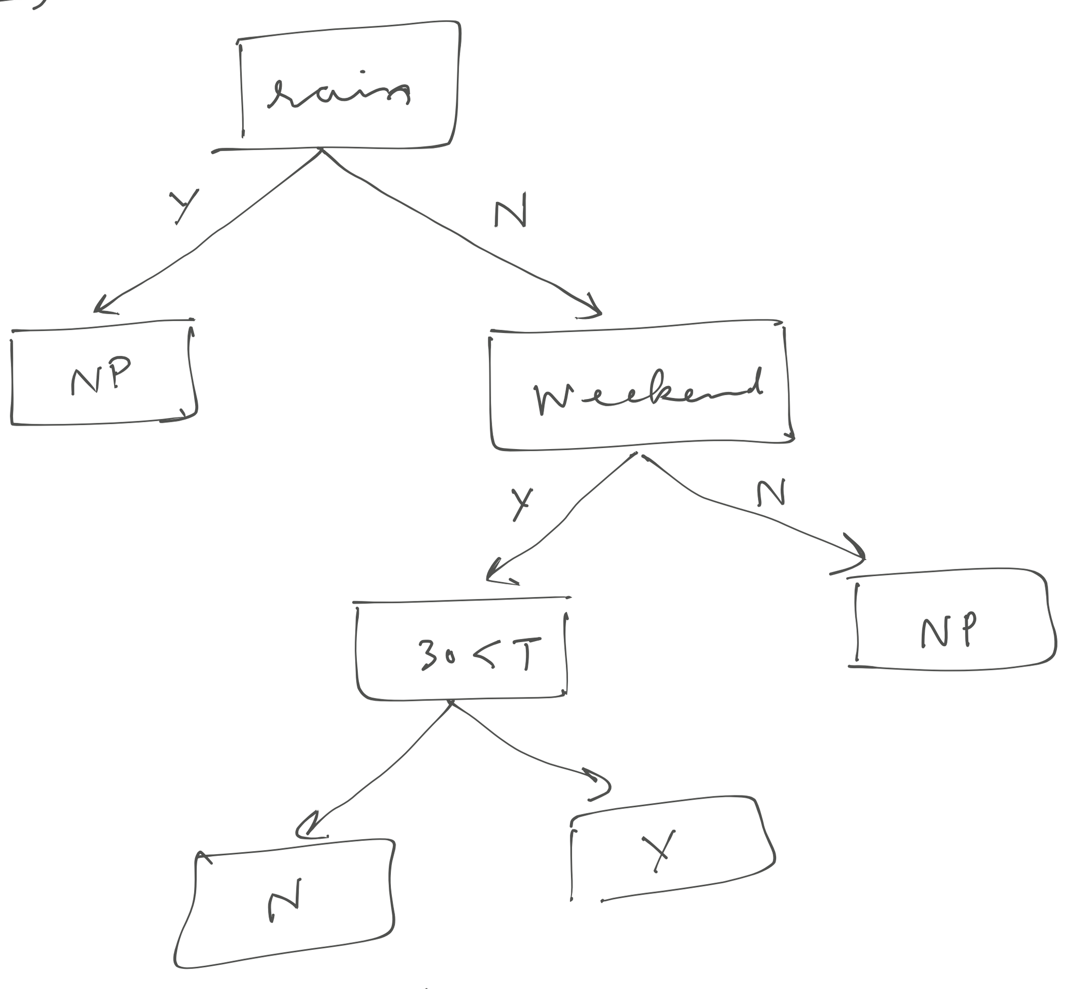
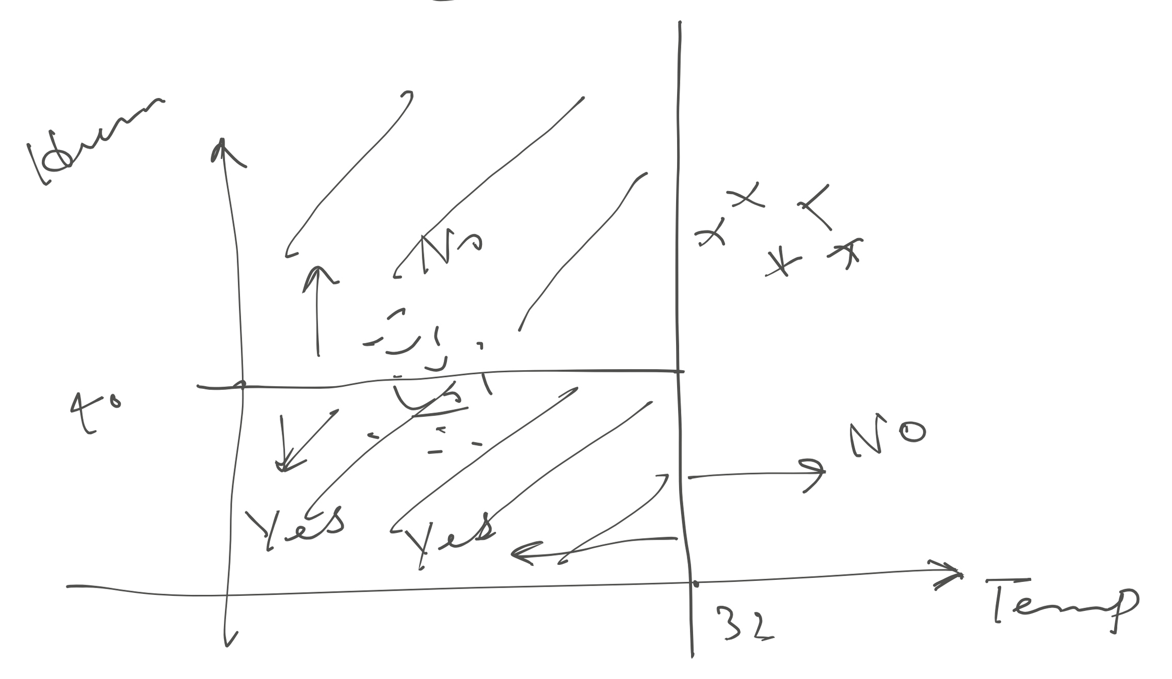
Decision Trees

rain → yes/no
Weekend → yes/no
Temp →

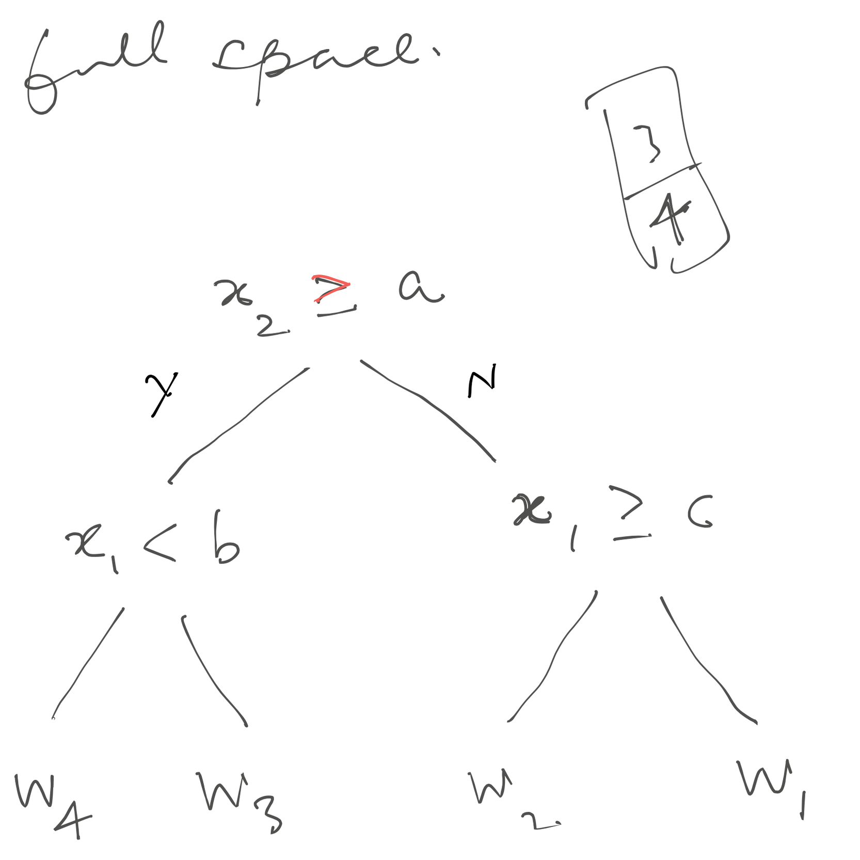


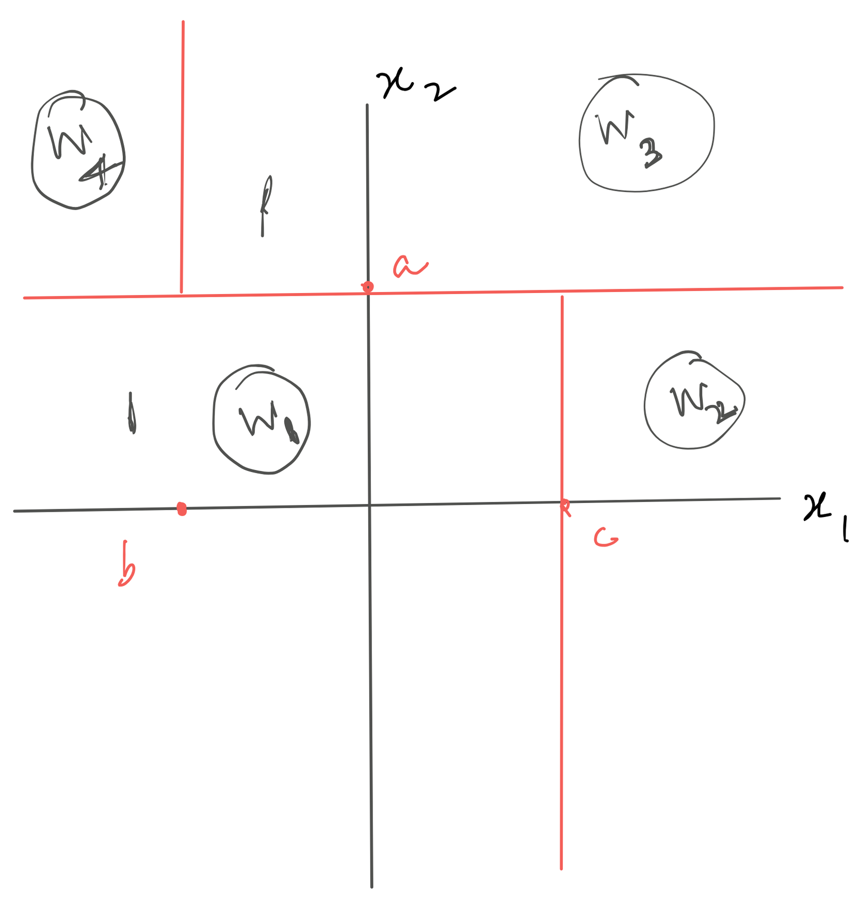


Partition your input space

There is no intersection between the partitions

Union of all the partitions returns the full space.





$a = 10$
$c = 15$
$b = -10$

Model fitting

Dataset (N, d)

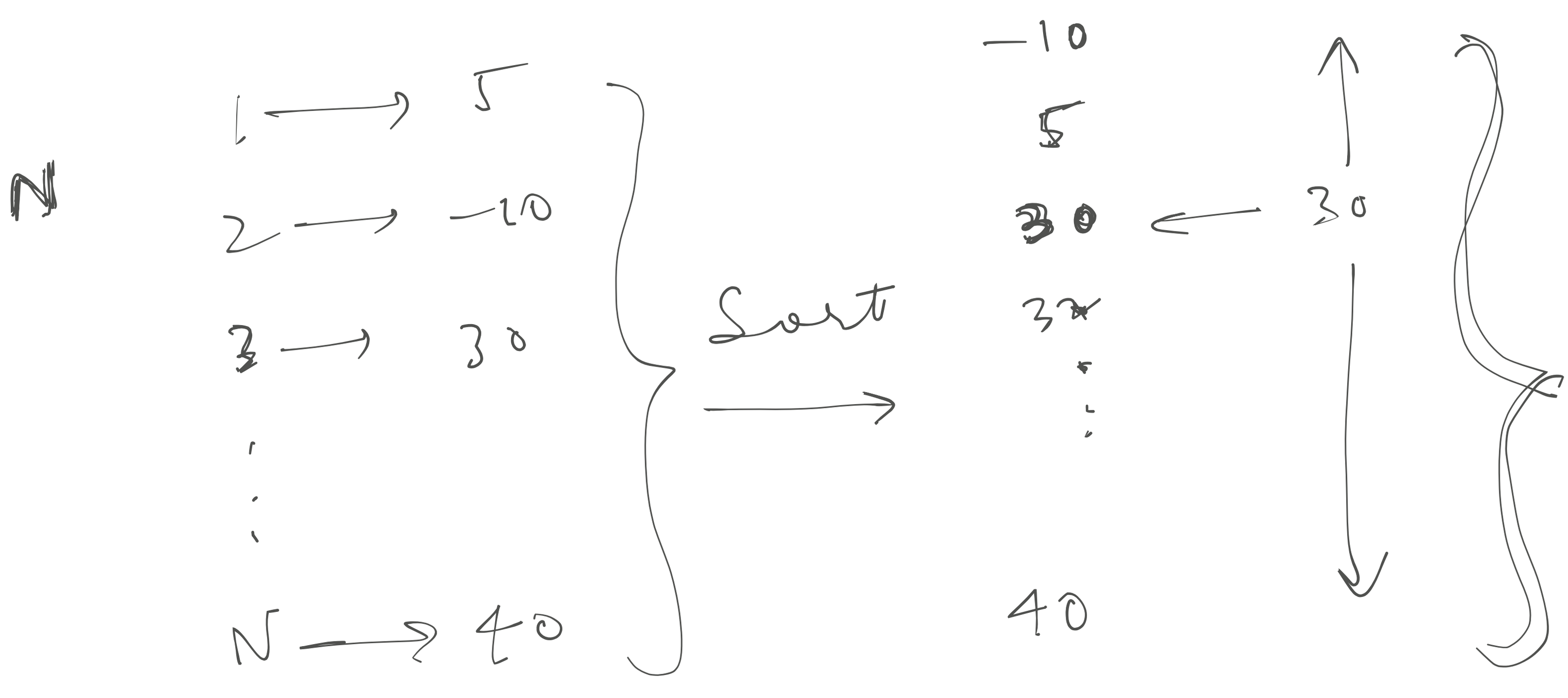
↳ Split quality

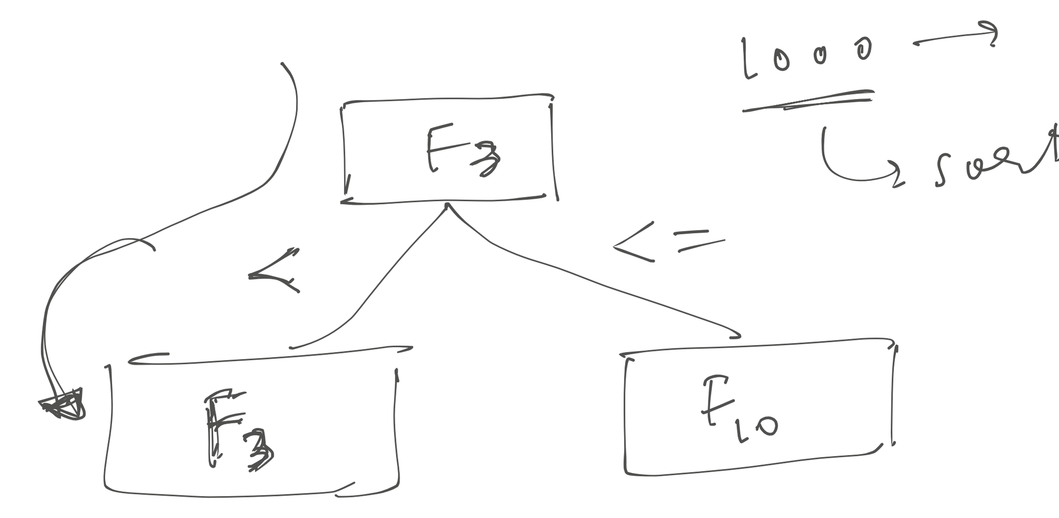
↳ Brute force → Try everything
↳ Greedy

Entropy

Select the best dimension

d → Temp   10 ___ 50

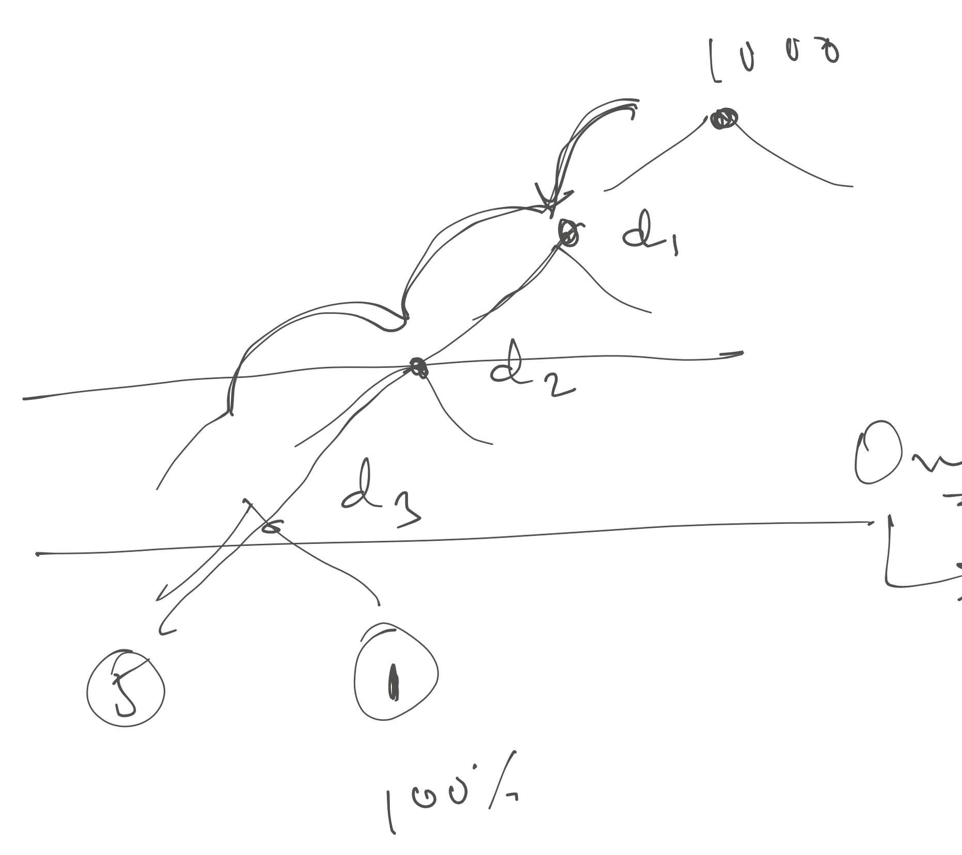




1000 → 10
↳ sort → E

Co 90% 10%

Decision → Majority class → Co

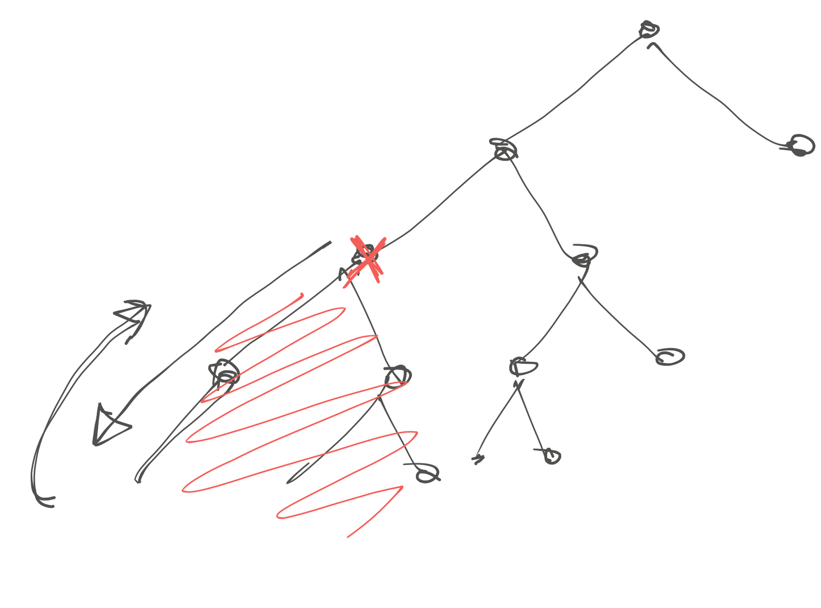


Overfitting
↳ Heuristics

Pruning
Pre
Post

Nodes < L = 50

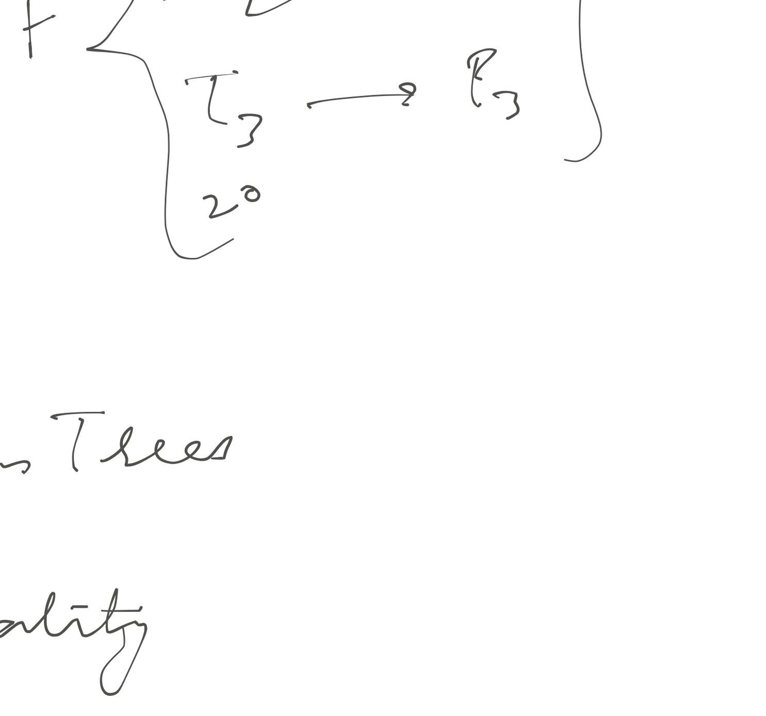
Depth < d = 3



Random Forest

1000

T_1 ← 800   200
T_2 ← 800   200
T_3 ← 800   200



✓ Overfitting in Decision Trees

✓ Entropy → Split quality
Gini Impurity
Information gain