Decision Tree - 2

Agenda

⇒ Recap

⇒ Gini impurity

⇒ Spliting on Numerical feature

⇒ Imbalanced Data

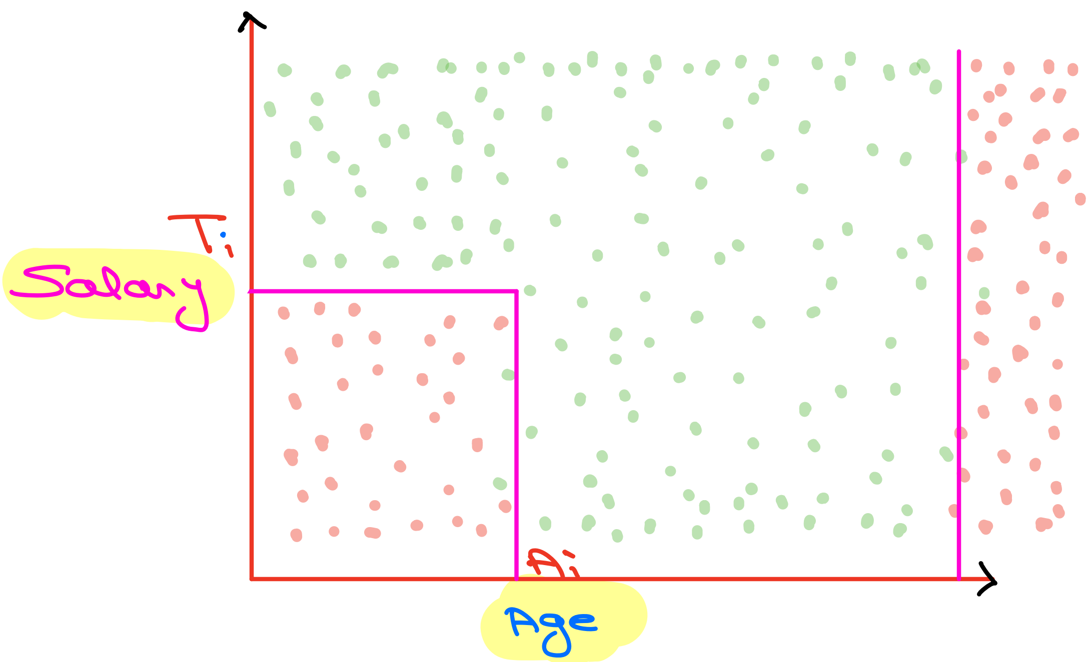
⇒ Feature Scaling

⇒ Feature importance

⇒ DT Regression

Recap

```
if salary < Ti:
    if age < Ai:
        P = +1
    else
        P = -1
else
    .......
```



① Decision tree splits Data into Homogeneous regions using Axis parallel Hyperplanes.

② DT is easily interpretable

⇒ How do we decide which feature and Value to Split On

⇒ Target: Pure/ Homogenous Node

⇒ How do we calculate purity?

$$\boxed{\text{Entropy}(H(Y)) = -\sum_{i=1}^{k} P(y_i) \log P(y_i)}$$

where k classes

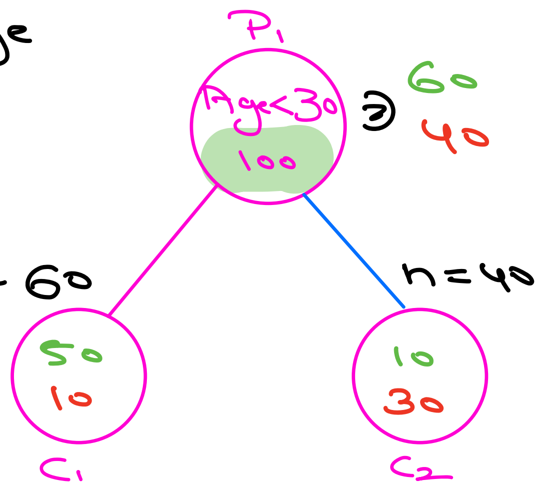
⇒ For Binary class : k=2

P or 1−P

$$H(Y) = -\left(P \times \log P + (1-P) \times \log(1-P)\right)$$

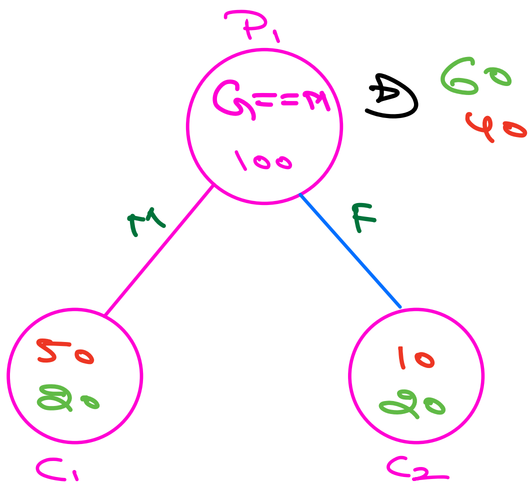
F1 ⇒ Age

Ex :



$H(C_1) \Rightarrow 0.65$ $H(C_2) \Rightarrow 0.81$

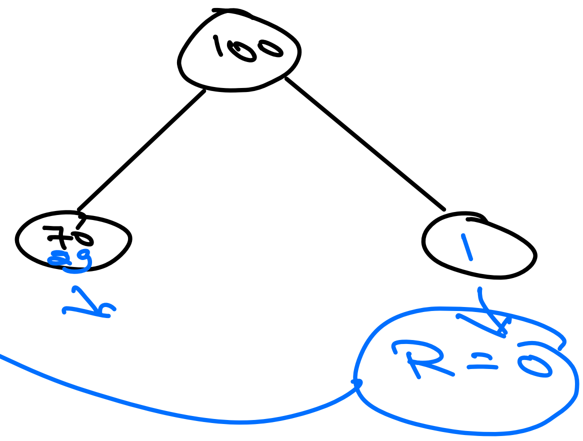
F2 = Gender



$H(C_1) \Rightarrow 0.86$ $H(C_2) = 0.91$

Total Entropy for Each Split

why not : $\frac{C_1 + C_2}{2}$



$$\frac{n_1}{n_{Total}} \times H(C_1) + \frac{n_2}{n_{Total}} \times H(C_2)$$

$F_1 \Rightarrow 0.81$ $F_2 \Rightarrow 0.67$

Information Gain
(Reduction in Entropy)

$H(Parent) - H(Child)$

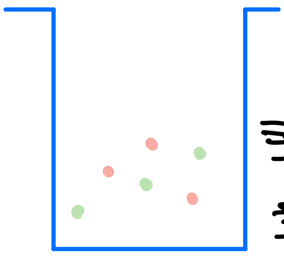
Limitation of Entropy

$\Rightarrow$ Lot of Calculations involve log

Gini Impurity

$$GI(y) \Rightarrow 1 - \sum_{i=1}^{k} (P(y_i))^2$$

H(y) | GI(y)



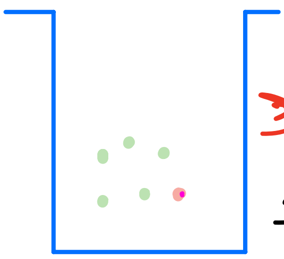
$\Rightarrow 0.5 \times \log(0.5) + 0.5 \log(0.5)$

$\Rightarrow 1$

$1 - ((0.5)^2 + (0.5)^2)$

$1 - ((0.25) + 0.25)$

$\Rightarrow 1 - 0.5 \Rightarrow 0.5$

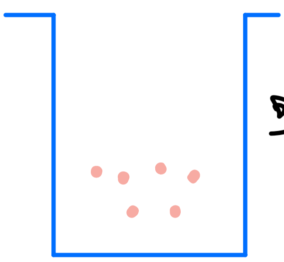


$\Rightarrow \frac{1}{6} \times \log \frac{1}{2} + \frac{5}{6} \times \log(\frac{5}{6})$

$\Rightarrow 0.6$

$1 - ((\frac{1}{6})^2 + (\frac{5}{6})^2)$

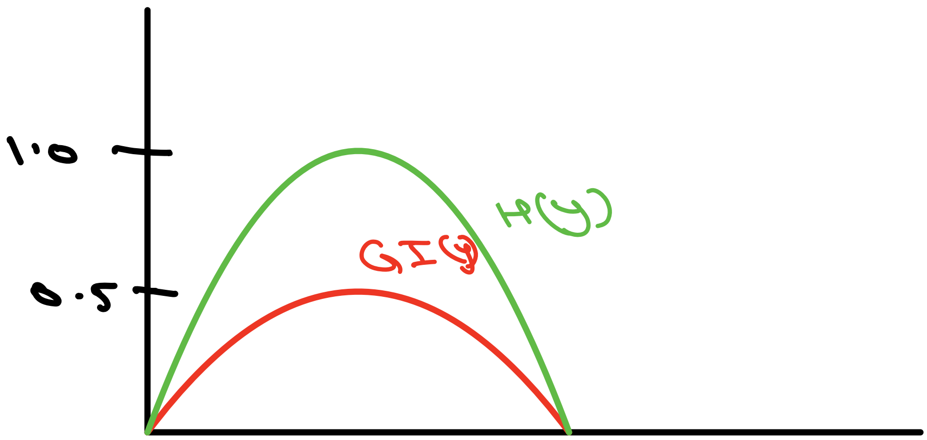
$\Rightarrow 0.27$



$\Rightarrow 0$

$\Rightarrow 1 - (1^2 + 0^2)$

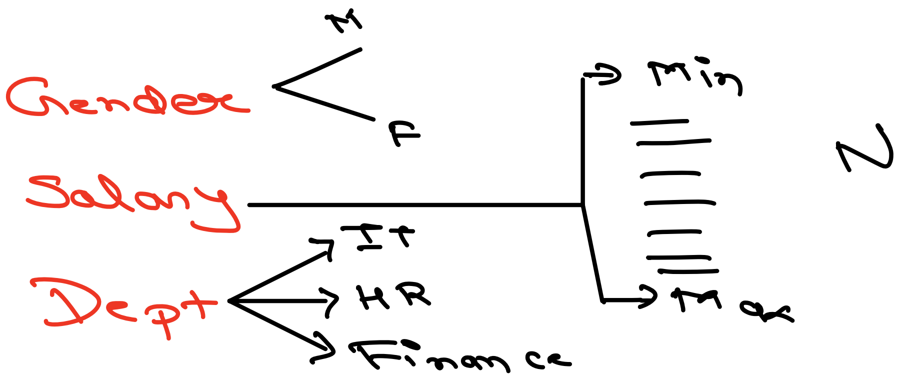
$\Rightarrow 0$



F_1 and F_2

$w(G(y_1))$ 0.32 $w(G(f_2))$ 0.24

Splitting on Numerical

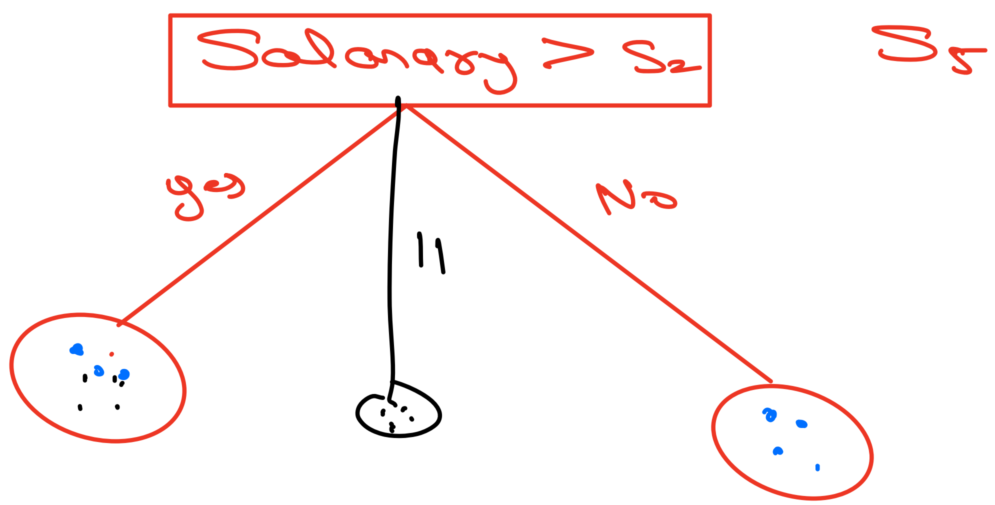


Brute Force

Step1: Get all Unique Values of Salary
$[S_1 \ S_2 \ S_3 \ --- \ S_n]$

Step2: Calculate IG for each Salary Value as threshold
$[IG_1, IG_2 \ IG_3 \ --- \ IG_n]$

Step 3: Argmax of IG



* Problem with this approach

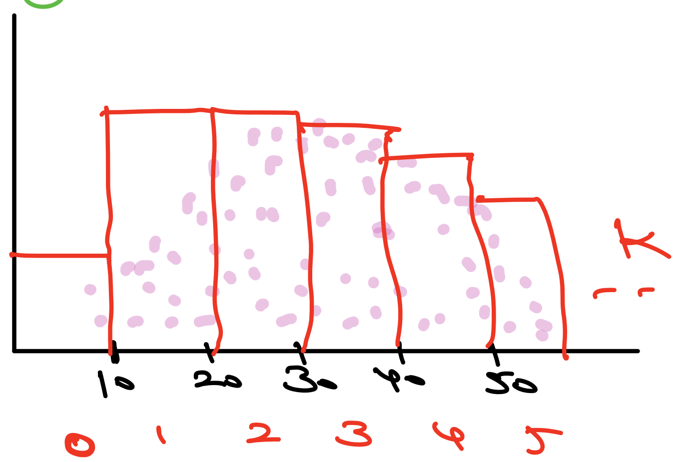
d numerical feature

n rows

$O(n \times d)$

How do we solve this

Binning



K

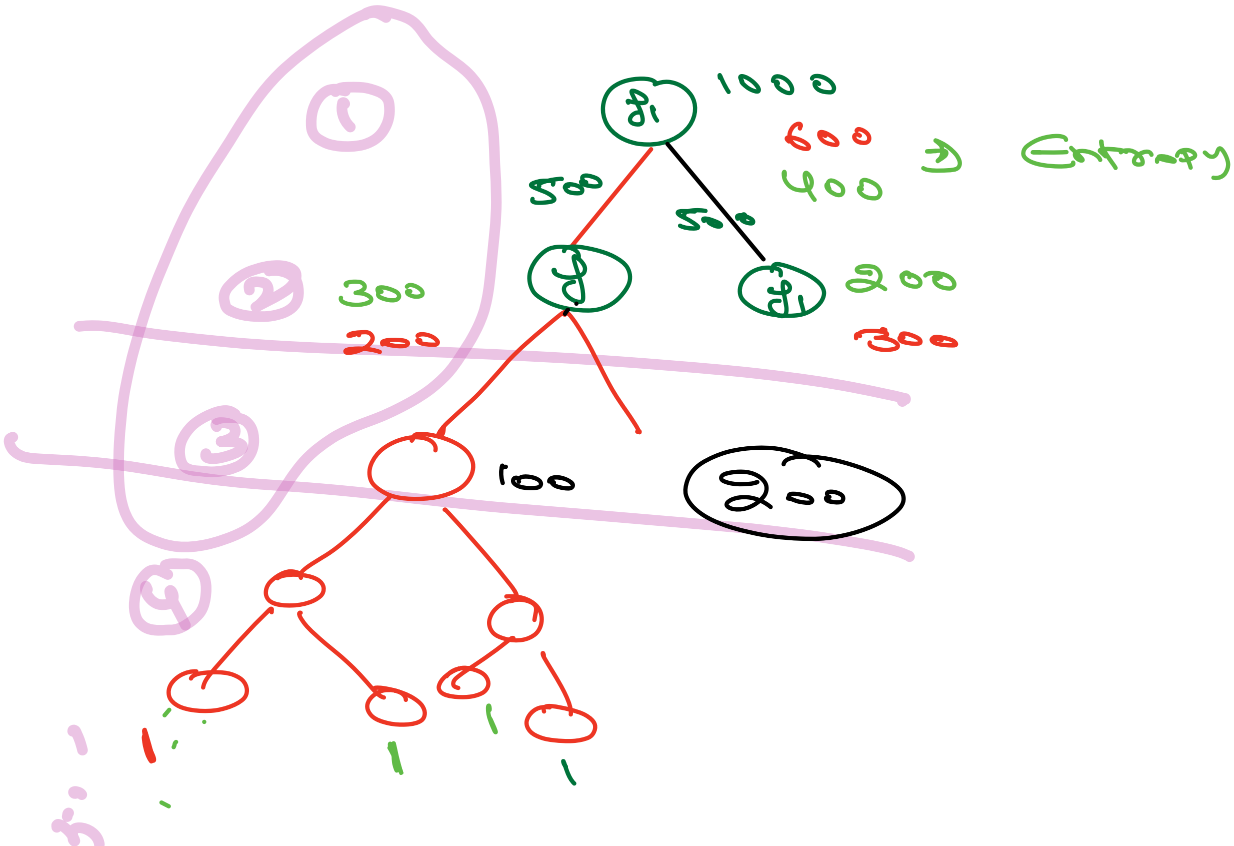
$n \le 10000$ Unique Salary $\Rightarrow$ n entropy

↓

Binning

↓

k entropy



Total accuracy = 100% Train Acc.

Decision Tree is very to Overfitting

max-depth → 5

* Pruning: Cutting Un-necessary Branches

① Max-depth

② Min-samples-split : Min nodes for Split Consider

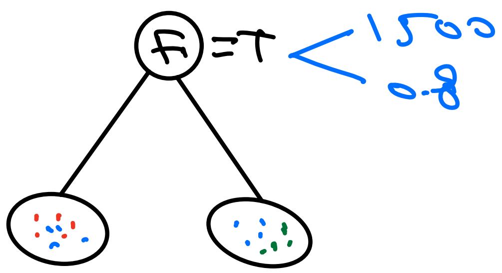
③ Min-samples-leaf → Min dp every Node Should have

④ Max-leaf-Nodes → Max number of Nodes Tree Can have

Q- Will feature scaling have any impact on DT?

f1	f1-scaled
2000	1.5
2500	1.8
1500	0.8

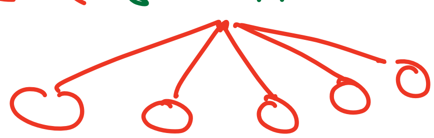
No impact



No need to Scale for DT

Q Categorical feature

⇒ 5 Unique Categorical : OHE



⇒ 10000 Unique Categories

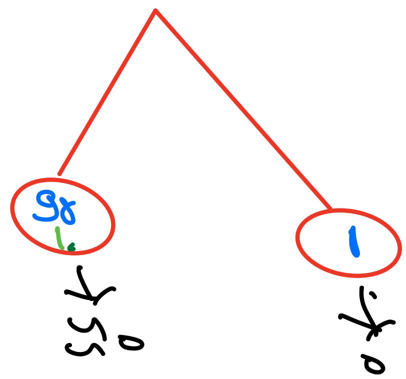
Target Encoding

Q Impact of Outliers ⇒ Yes

↳ prune Tree

Q Data Imbalance Impact DT ⇒

100 → 99
→ 1



Entropy ≈ 0

$1 \Rightarrow \frac{98}{99}$ %

How do we solve this

↳ Class-Weights

↳ SMOTE

↳ Undersampling / Oversampling

for Tmrw

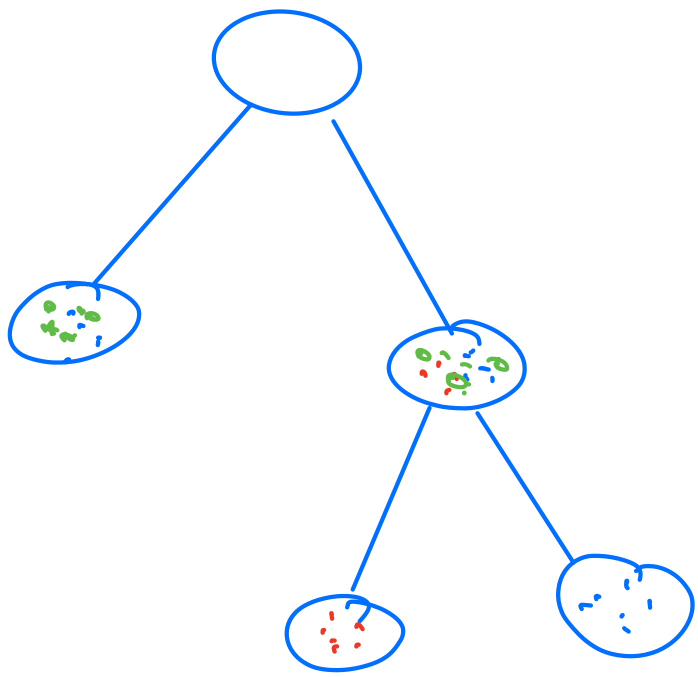
Feature Importance

Decision Tree for Regression

Smote ⇒ To create Artificial D.P

k - points

$$P'_1 = \lambda_1 P_1 + \lambda_2 P_2 + \lambda_3 P_3$$

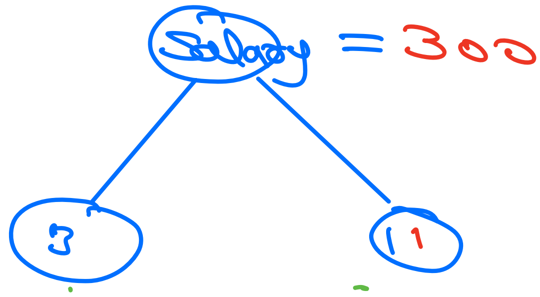


min_leaf

max_dep

1,1	1,2	1,3
2,1	2,2	2,3
10,1	10,2	10,3

Salary	Churn
5000	1
3000	1
2000	0
5000	1
1000	1



Salary	IG
5000	x_i
3000	x_2
2000	x_3 → IG is Max
	1
1000	