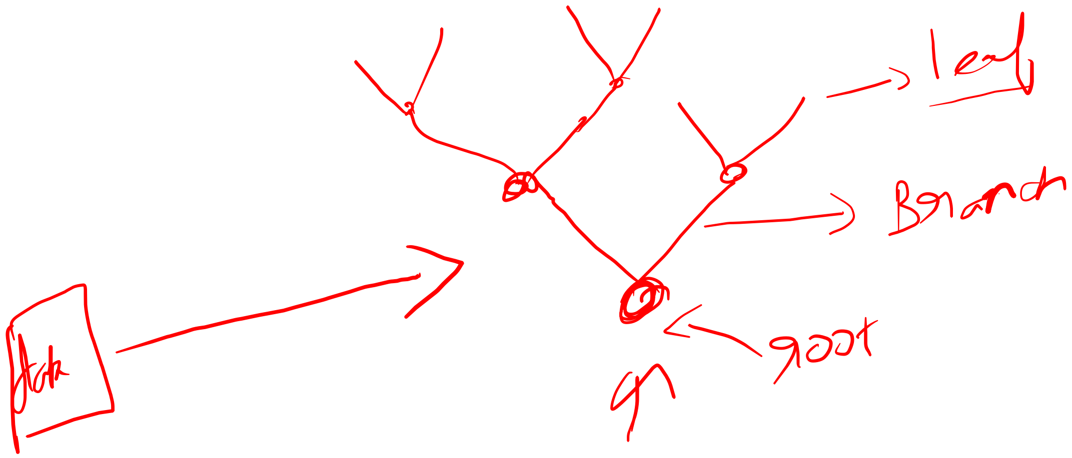
Tree

DS

→ Tree

prediction

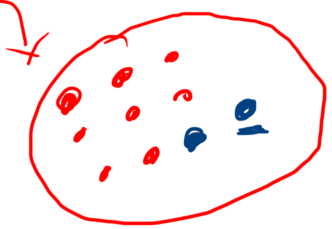
?



2 How this decision will be prepared ?

10 R.

1



what is the error → 1 blue

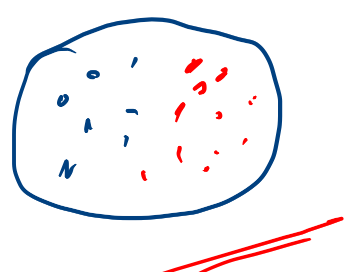
9 green → information about red → 0.9

8 red → 0.8 →

↳ 0.2 → error.

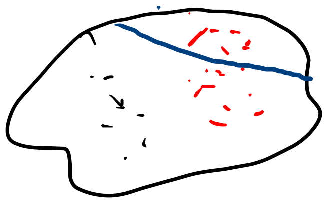
2 Pure Set & Impure set

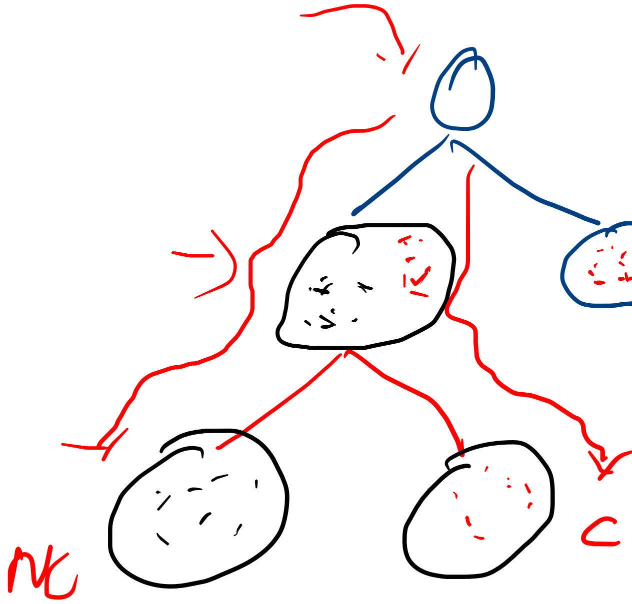
13



entropy $= \sum -(p_i)\log_2(p_i)$

How pure is the given dataset



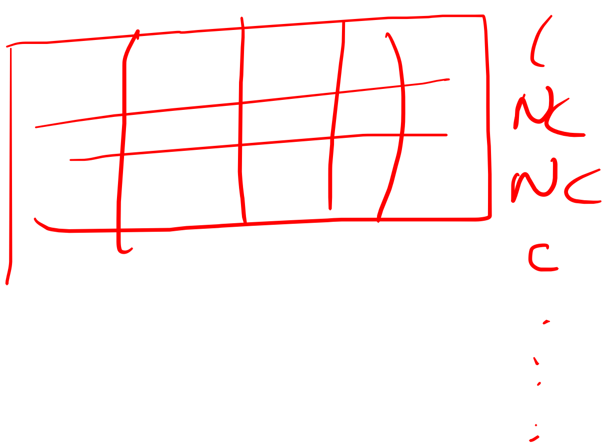


← impurity

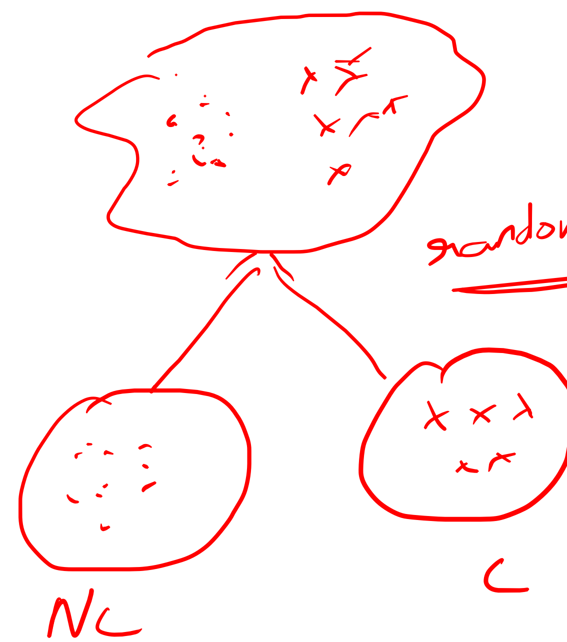
decision Tree

until it get pure set

It will be splitting data



mixed group



random split → impure

yes

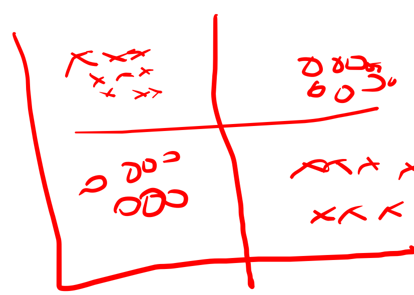
100 patient

build this line using logistic regression

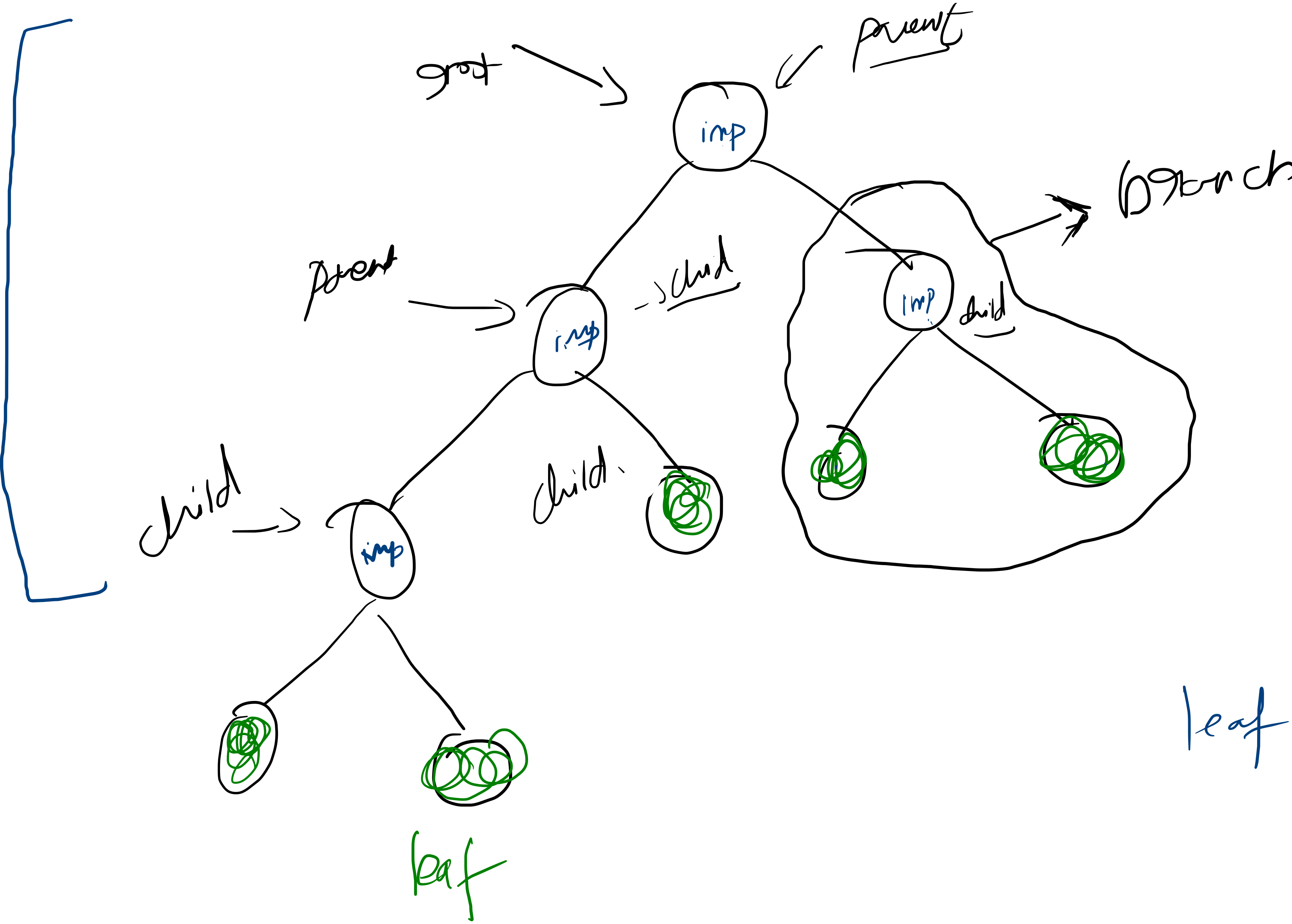
knn, decision tree

1) more than 2 labels

2) linear seperable vs non linear seperable

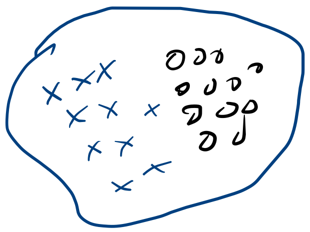


knn

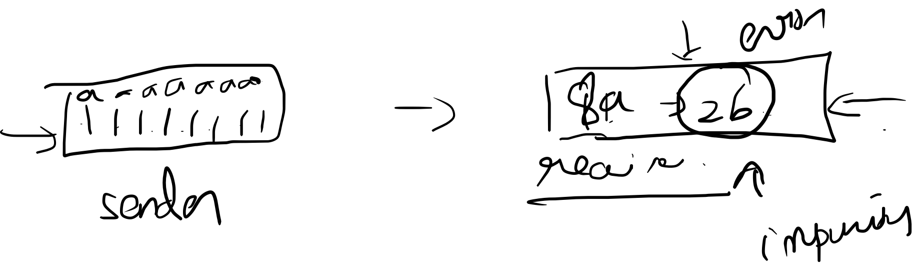


leaf - can only pure set

prediction



→ Entropy



5B 5R

8B 2R

B → 0.8

R → 0.2

1. entropy
2. chi square
3. gini

$$\text{Entropy} = \sum -p_i \log_2(p_i)$$

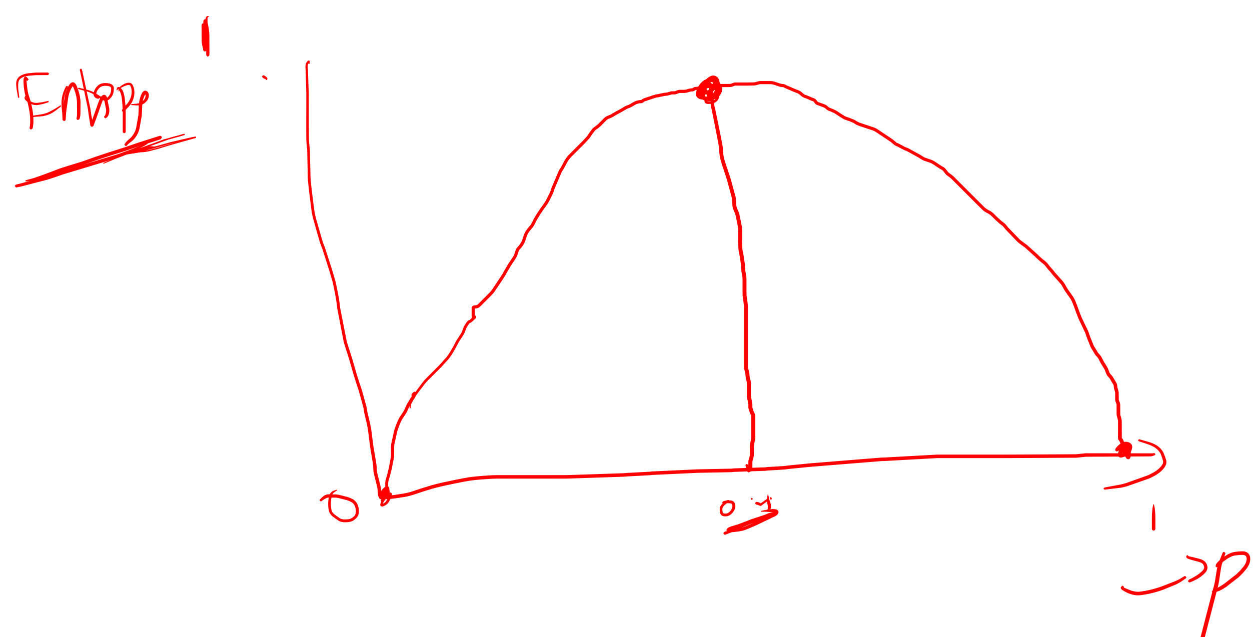
$$= -0.8 \log_2(0.8) - 0.2 \log_2(0.2)$$

$$= 0.2$$

10 R

9R 1B

8R 2B

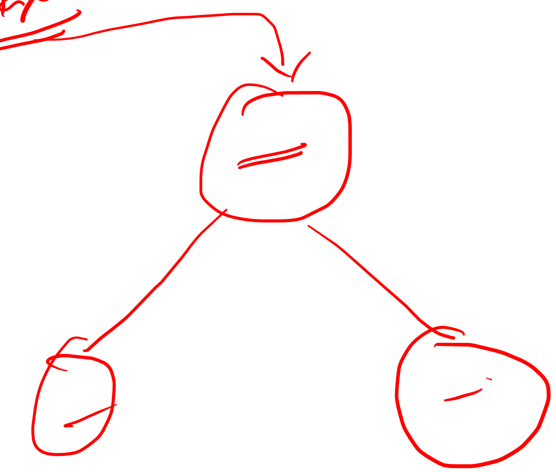


Information gain $= 1 - \boxed{\text{Entropy}}$ ← of system

$= 1 - 0.2$

$= 0.8$

Max



random split → entropy → information gain → pure

not pure → random split

$$E(T, X) = \sum P(c) * E(c)$$

$=$

$\underline{\text{Entropy}} = \underline{0.94}$

$\text{Entropy} = \Sigma - p_i \log_2 p$

$\Sigma\ \underline{p(c)} * E(c)$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

$5/14\ E(3,2) + 4/14\ E(4,0)$

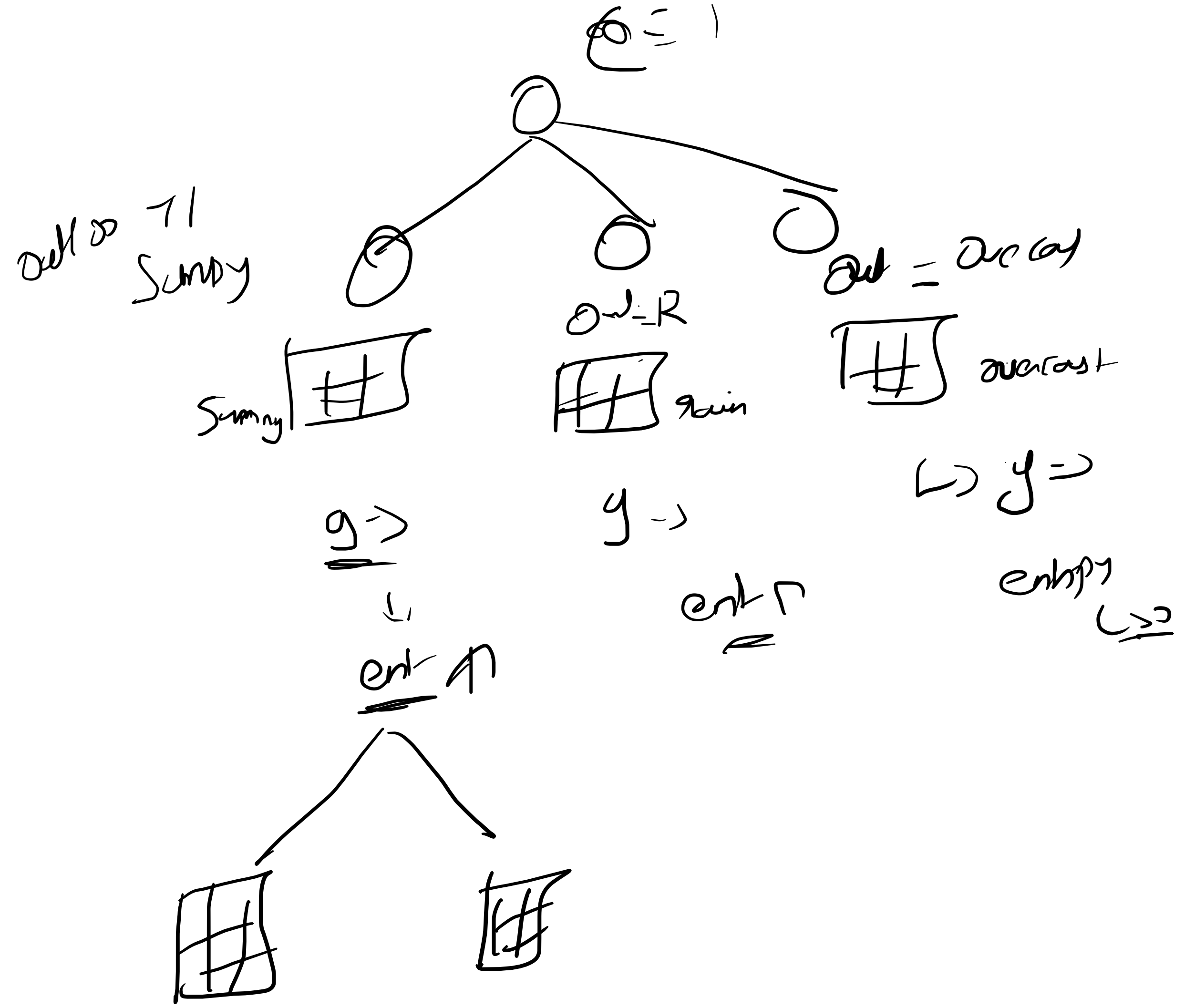
$+ 5/14\ (2,3)$

$= \underline{0.671}$

$E(T) \qquad E(T,X)$

Information gain of outlook $= \underline{0.94} - \boxed{0.671}$

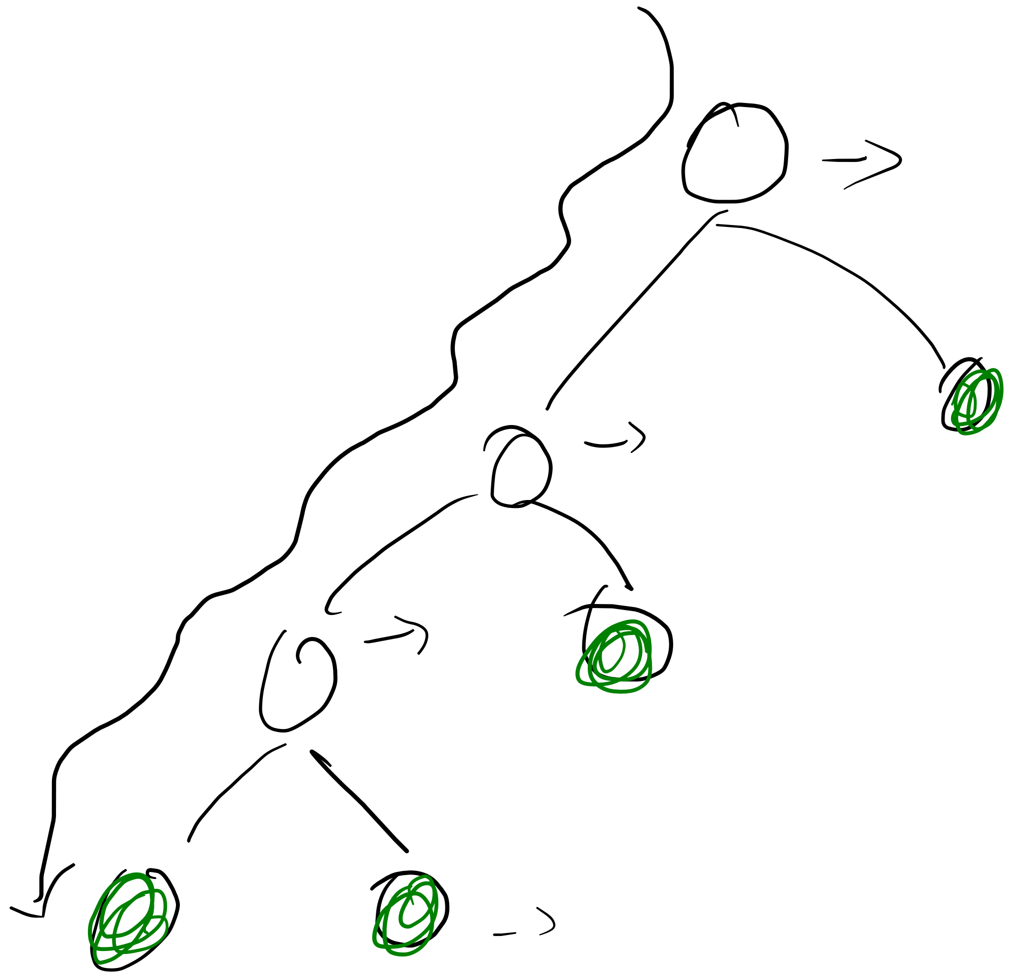
$= \underline{0.264}$



1. entropy T
2. entropy T, X
3. gain

depth of Tree

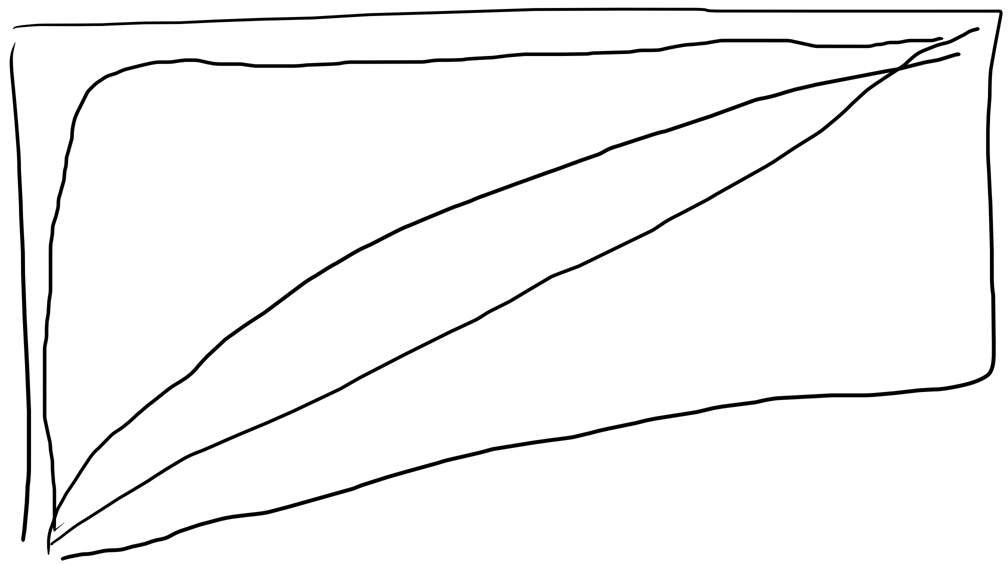
4



$$gini = 1 - (p^2 + q^2)$$

$$g' = 1 - \sum p^2$$

$$g_{(T,x)} = 1 - \sum p^2_{(i/t)}$$



$$gini_{(p,s)} = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = 0.2$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

$$gini_{(p,o)} = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2 = 0$$

$$gini_{(p,r)} = 1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2$$

$$= 0.3$$

$$gini\ interest_{(Sunny)} = \frac{5}{14} \times 0.2 + \ldots$$

$$1 - \ldots$$

$$= 0.624$$

$$gini\ gain = gini(p) - gini(outlook)$$