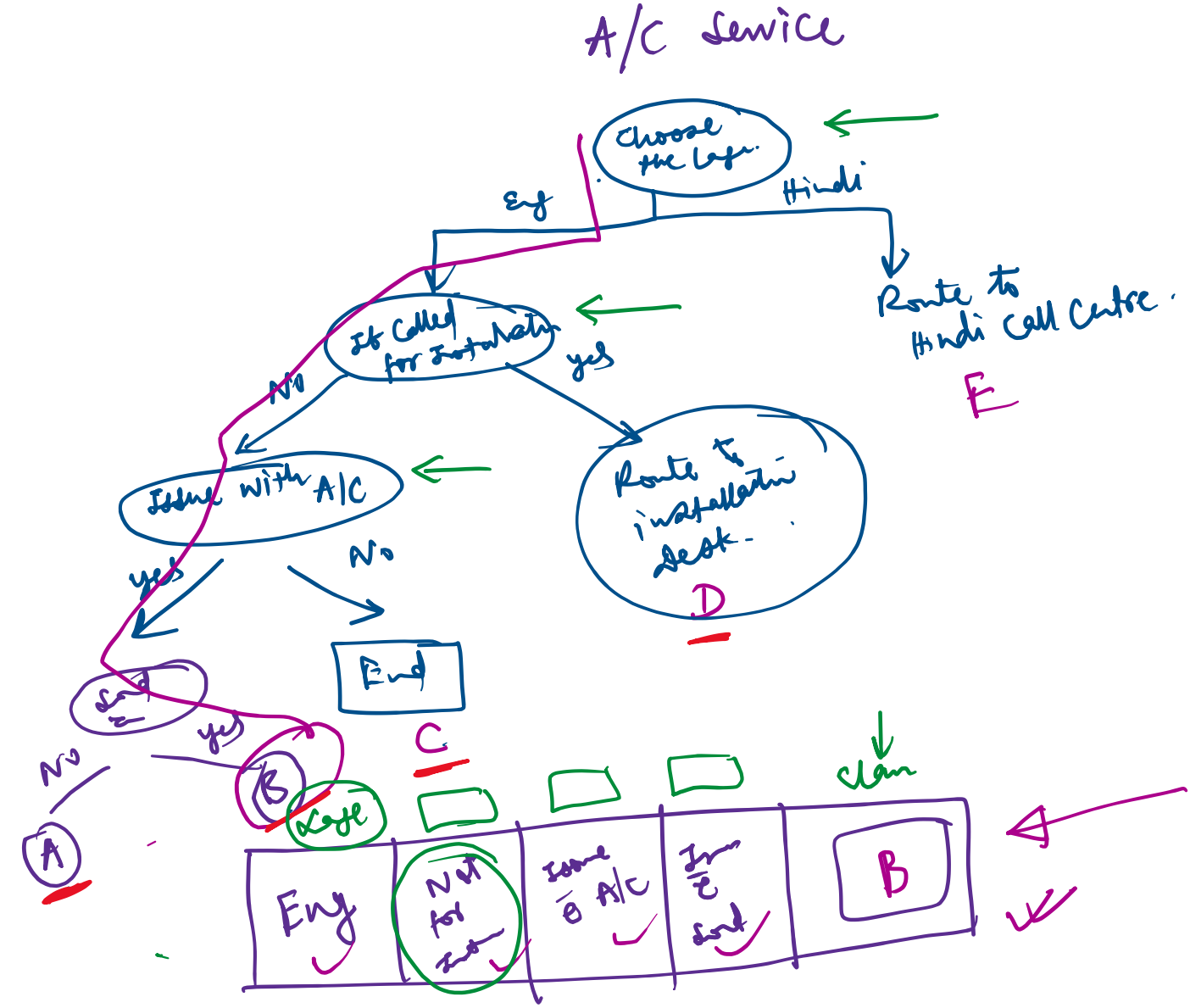
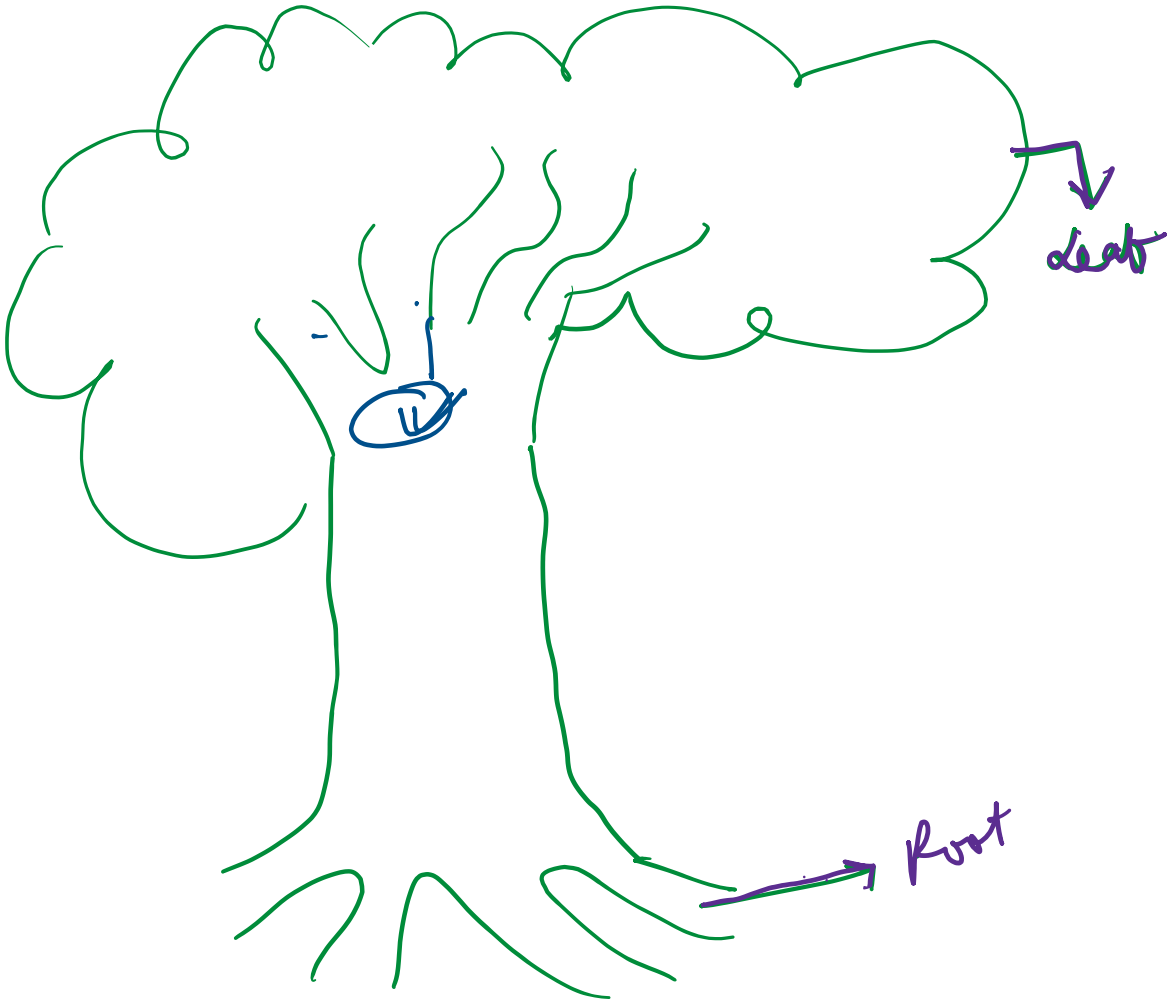
Decision Tree → Tree based structure

Based on decision.

→ Supervised ML Algo.
→ Mostly used for classification but can be used in Regression.

→ Tree based structure.
→ Based on decisions.





Terminologies:

1. Root node: Main Node (splitting starts here).
2. Splitting: Process of dividing Nodes into Subnodes.
3. Decision Node: Condition node, at the node where (decision is made / Nodes ↓ Subnodes)
4. Parent/Child Nodes: Node which is divided into sub nodes are called Parent Node of Sub-nodes
   — o —
   These sub nodes are the child of the Parent node.

(5) Subnode = Child node.

(6) Terminal Node / Leaf Node :- Last Node (No further split).

(7) Pruning : Process of Removing subnodes
↳ (Reverse of splitting)

DT

→ Algorithm

splitting criteria

① CART → uses Gini

② C4.5 → " Entropy

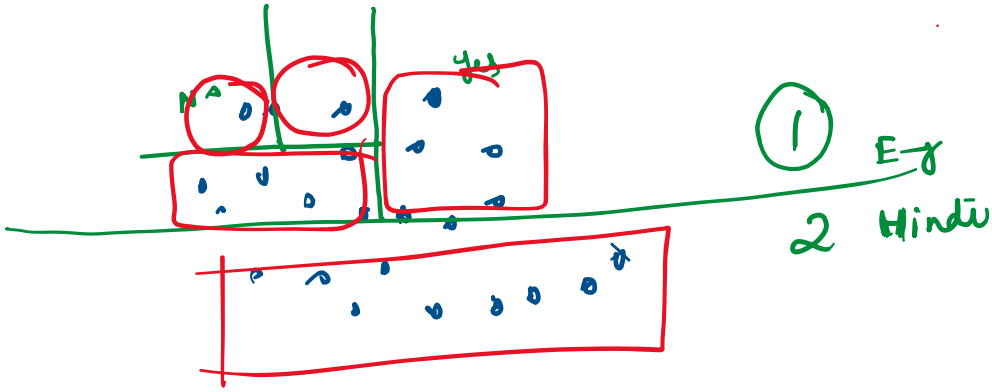
③ ID3 → " "

④ CHAID → Chi Square

→ Based on Recursive Partitioning

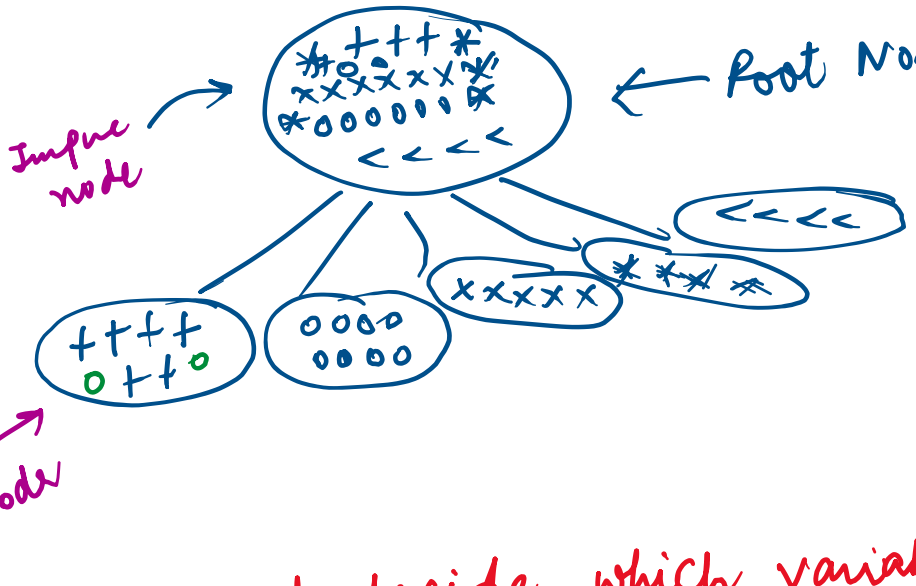
→ Hierarchical

→ Based on different splitting criteria.



splitting criteria

The Purpose is to get more pure nodes further.



Q. In D.T. how to decide which variable is used for splitting ?

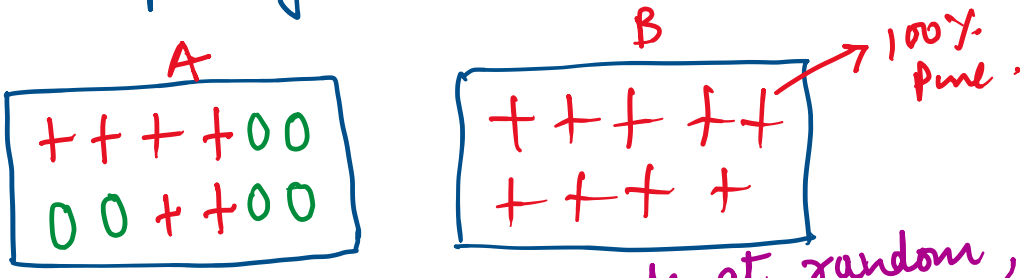
→ Based on splitting criteria

(1) gini (2) Entropy (3) Chi-Sq. (4) Red. in variance

Central Idea: To get more/better pure Node.

① Gini Impurity

Gini Impurity = 1 − Gini



→ If we select 2 items from a node at random, they must be from the same class.

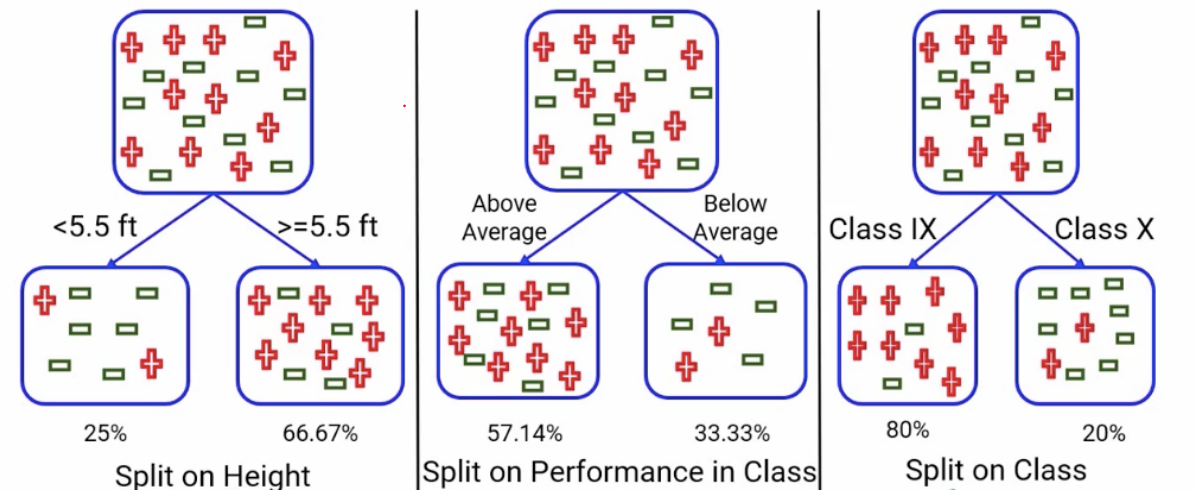
→ If it's 100% Pure — Prob. of Picking both items belonging to same class = 1 or 100%.

Steps:

① Gini Impurity of subnodes

② Gini = $(P_1^2 + P_2^2 + \cdots + P_n^2)$

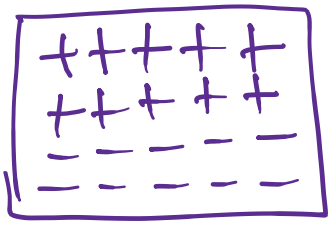
③ Gini Impurity of split
↳ taking the weighted gini impurity of subnodes of split.



height	Performance in Class	class	y = play / N.P.C.
<5.5	Above	IX	+
>5.5	below	X	−

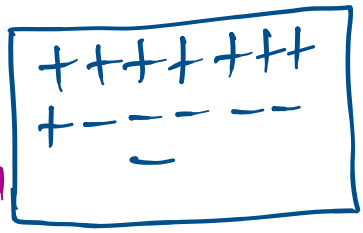
Split1: Split on performance

Student = 20
P.C. = 10
% P.C. = 50%
NPC = 10
% NPC = 50%

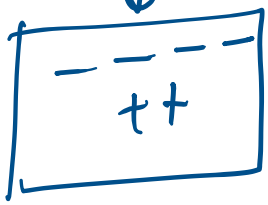


14/20 A.A.

6/20 B.A.



St = 14
P.C. = 8
N.P.C = 6
P(P.C.) = 8/14 = 0.57
P(NPC) = 6/14 = 0.43



St = 6
PC = 2
P(PC) = 2/6 = 0.336
P(NPC) = 4/6 = 0.67

(1) G.I. of Subnode A.A.

$= 1 - ((0.57)^2 + (0.43)^2) =$ **0.49**

(2) G.I. of Subnode B.A.

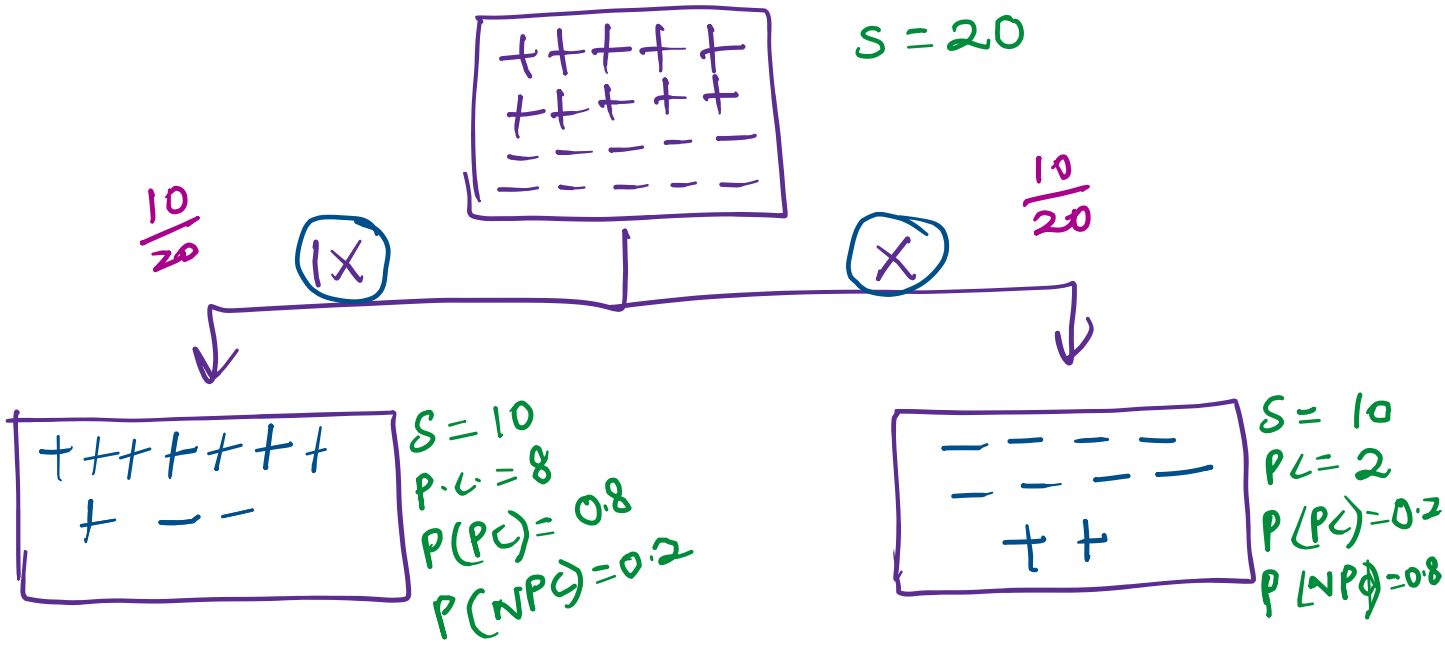
$= 1 - ((0.33)^2 + (0.67)^2) =$ **0.44**

(3) weighted G.I. of split

$= \frac{14}{20} \times (0.49) + \frac{6}{20} \times (0.44)$

$= 0.475$

split 2: split on class



① G.I. of subnode IX :

$$= 1 - [(0.8)^2 + (0.2)^2] = 0.32$$

② G.I. of subnode X

$$= 1 - [(0.2)^2 + (0.8)^2] = 0.32$$

③ weighted G.I. of split by class

$$= \frac{10}{20} \times 0.32 + \frac{10}{20} \times 0.32$$

$$= 0.32$$

Decision

split	weighted G.I.
Pest	0.475
class	0.32 ←

Since the G.I. is least when we split by class

hence

1st split split will happen by class.

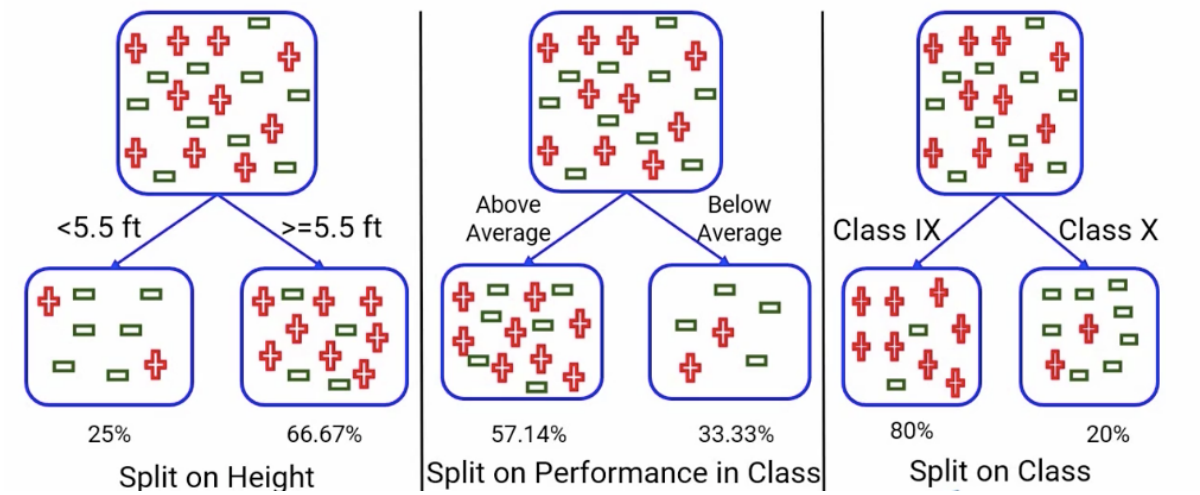
② Chi-Square.

→ Statistical significance of any node

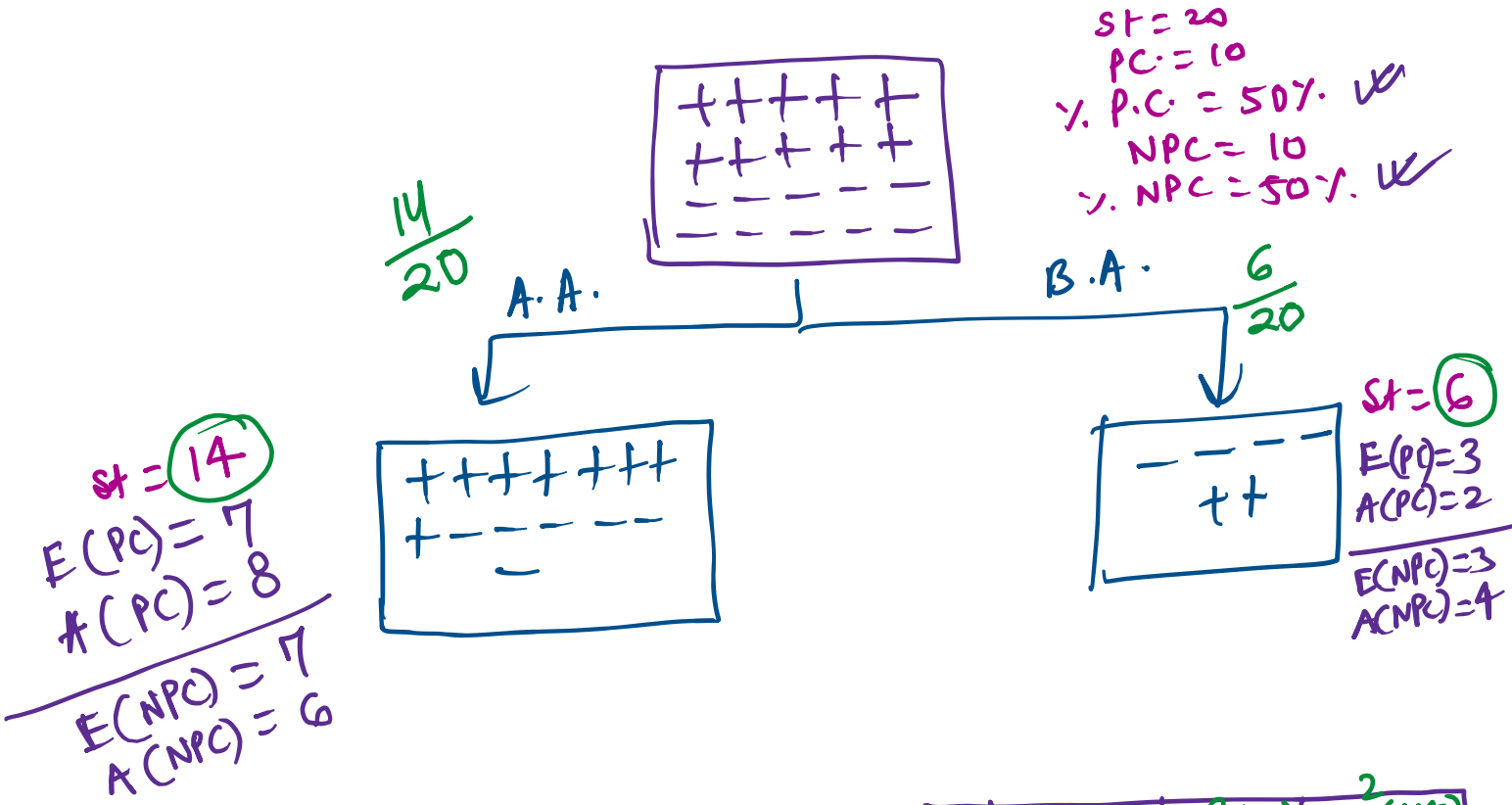
→ Mostly used for categorical var.

→ $\text{Chi-Sq}(\chi^2) = \sqrt{\frac{(\text{Actual} - \text{Expected})^2}{\text{Expected}}}$

→ Higher the value of χ^2, higher is the homogenity of node.



Split 1: Split on perf.

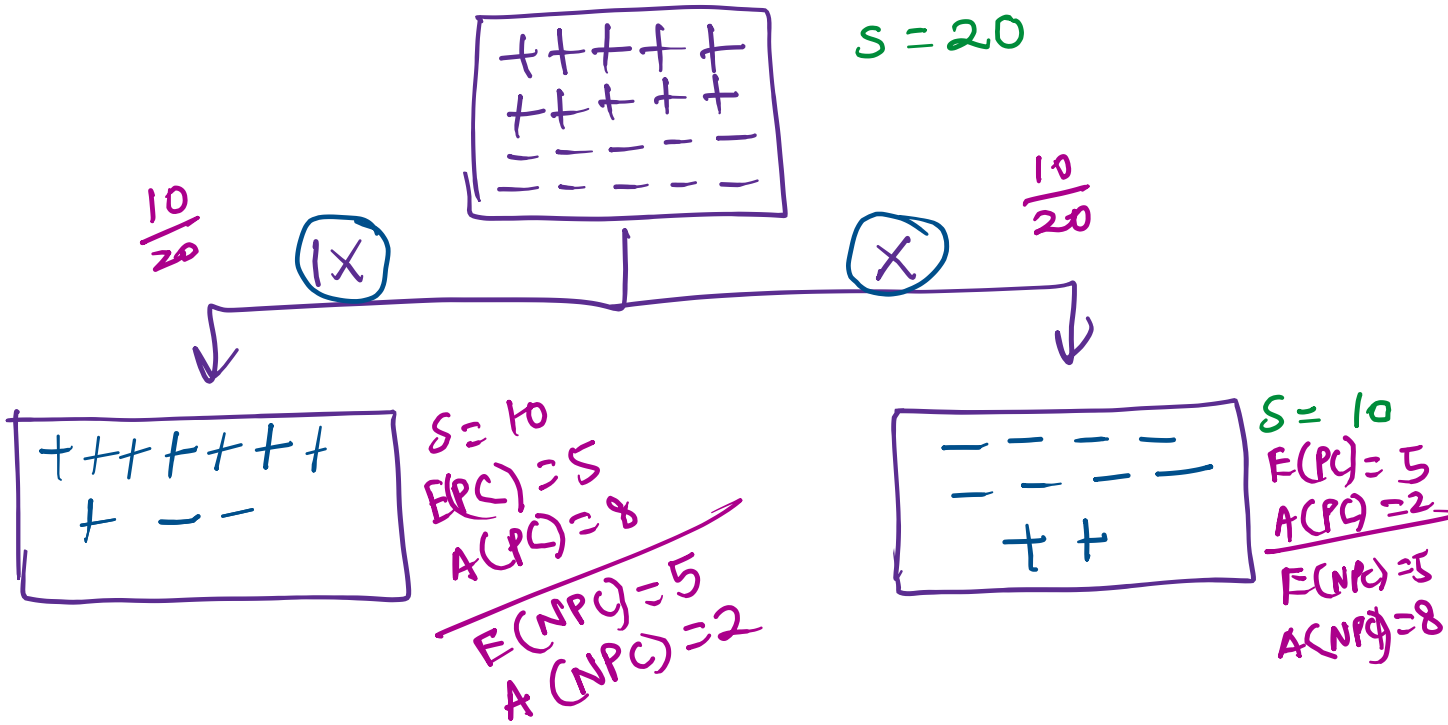


Node	A(P)	A(NPC)	E(P)	E(NPC)	Dev(PC)	Dev(NPC)	X^2(PC)	X^2(NPC)
AA	8	6	7	7	8-7 = 1	-1	0.37	0.37
BA	2	4	3	3	2-3 = -1	1	0.57	0.57

Total Chi-sq. for this split = 0.37 + 0.37 + 0.57 + 0.57

= 1.88

split 2: split on class



Node	A(PC)	A(NPC)	E(PC)	E(NPC)	D(PC)	D(NPC)	χ^2(PC)	χ^2(NPC)
1X	8	2	5	5	3	−3	1.34	1.34
X	2	8	5	5	−3	3	1.34	1.34

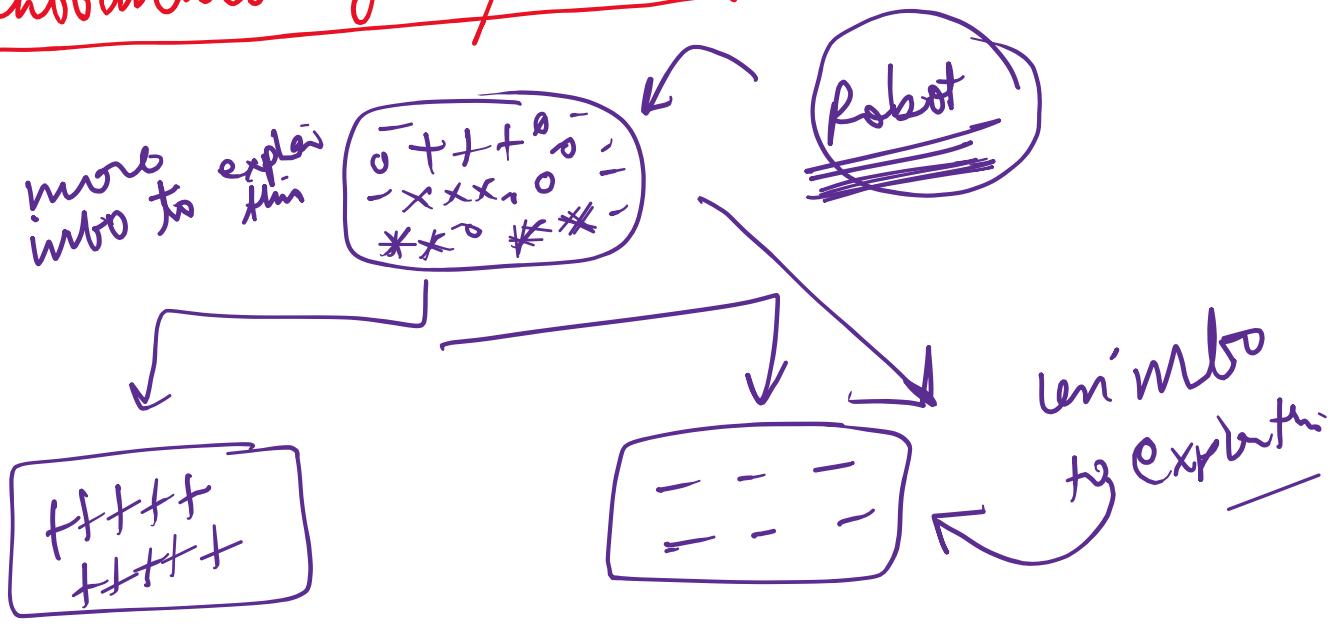
Total chi-sq. = 1.34 + 1.34 + 1.34 + 1.34

= 5.36

decision

split	chi-sq-
Pent	1.88
class	5.36

(3) Information gain / Entropy :-



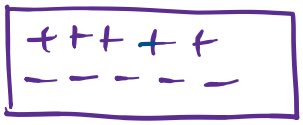
→ Higher information gain leads to more Pure nodes.

→ I.G. = 1 − Entropy

Entropy : $-P_1 \log_2 P_1 - P_2 \log_2 P_2 \cdots - P_n \log_2 P_n$

$\approx -P \log_2 P - q \log_2 q$

Ref:



A

P(+) = 0.5
P(−) = 0.5

B

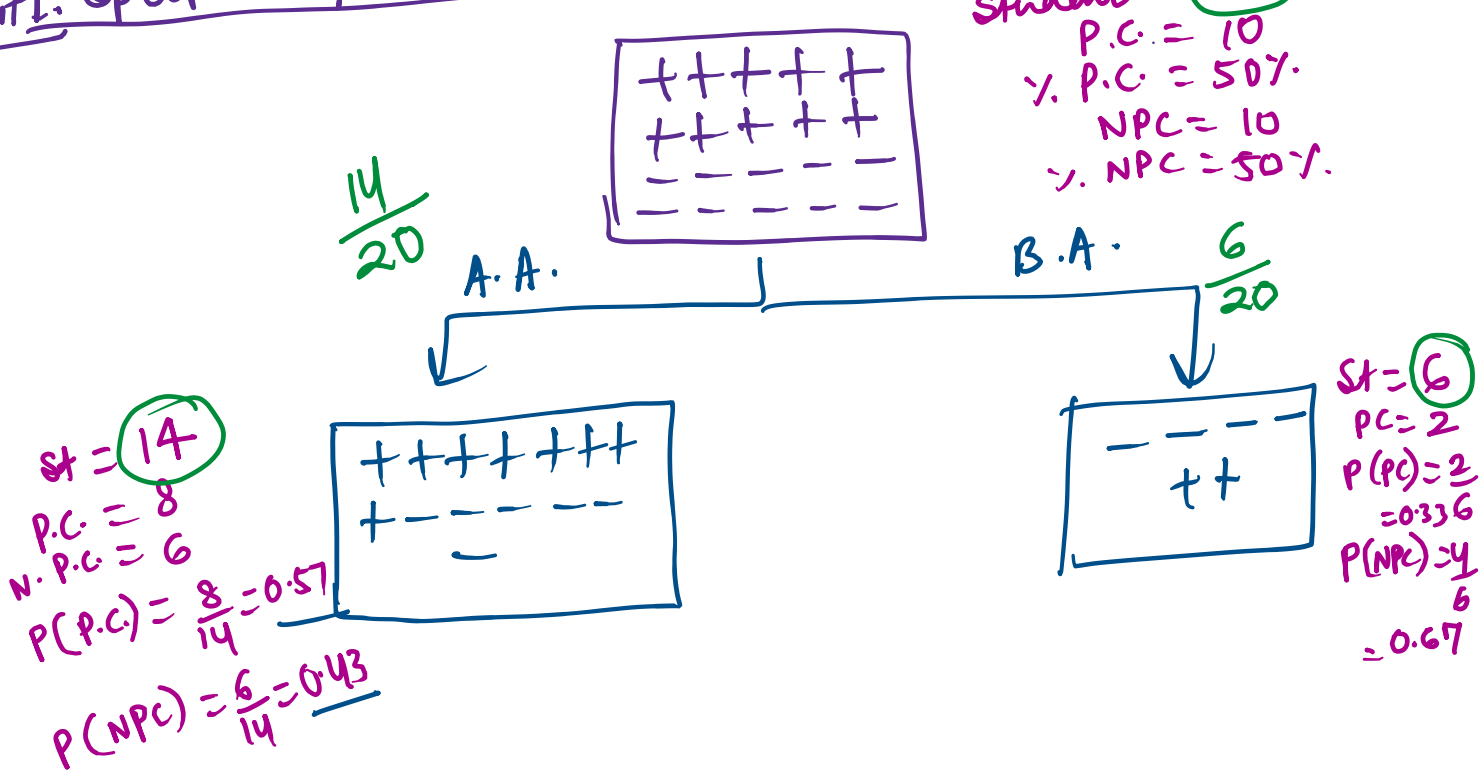
P(−) = 1
P(+) = 0

$$\text{Entropy}(A) = -P\log_2 P - q\log_2 q$$
$$= -0.5\log 0.5 - 0.5\log 0.5$$
$$= 1$$

$$\text{Entropy}(B) = -1\log 1 - 0\log 0$$
$$= 0$$

Higher Entropy ⟶ less Pure node

Split1: Split on performance



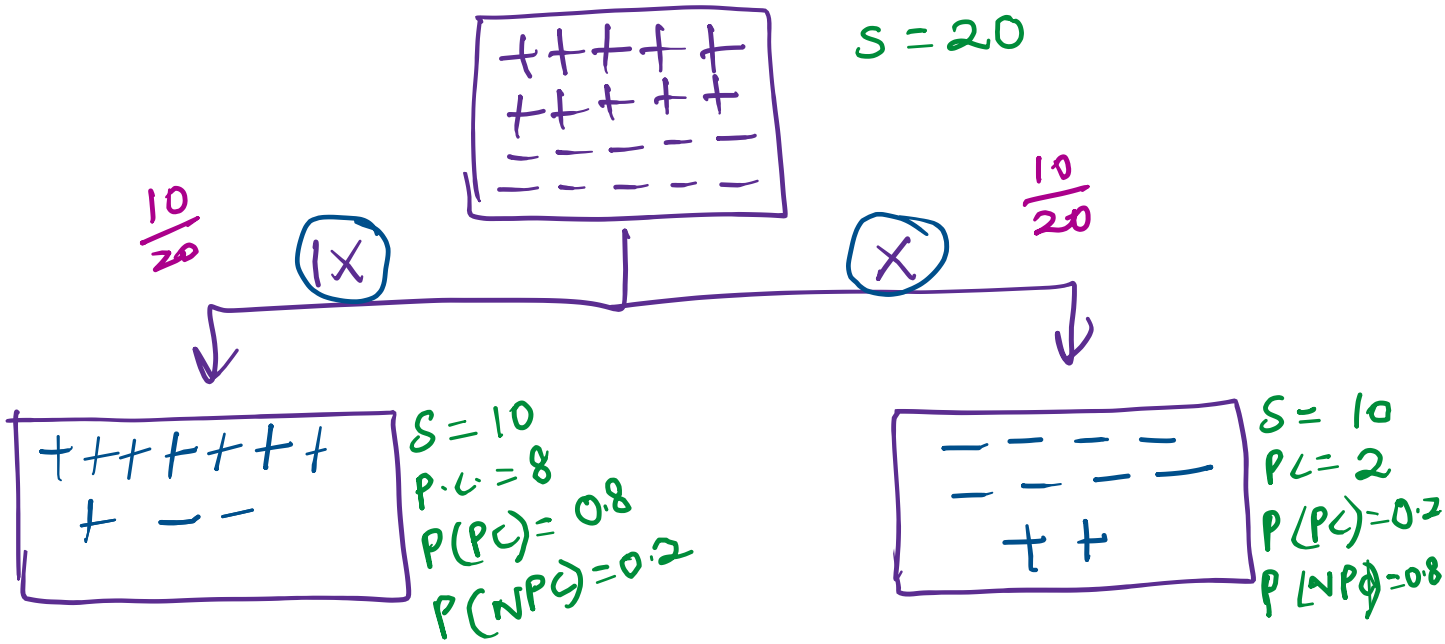
① $E(\text{Parent}) = -0.5 \log 0.5 - 0.5 \log 0.5 = 1$

② $E(\text{Subnode} = AA) = -0.57 \log 0.57 - 0.43 \log 0.43$
$= 0.98$

③ $E(\text{Subnode} = BA) = -0.33 \log 0.33 - 0.67 \log 0.67$
$= 0.91$

④ $\text{weighted entropy} = \frac{14}{20} \times (0.98) + \frac{6}{20} \times (0.91)$
$= 0.95$

split 2: split on class



① E(Parent) = 1

② E(1X) = $-0.8 \log 0.8 - 0.2 \log 0.2$
= 0.72

③ E(X) = $-0.2 \log 0.2 - 0.8 \log 0.8$
= 0.72

④ weighted Entropy for this split (by Class)

= $\frac{10}{20} \times 0.72 + \frac{10}{20} \times 0.72$

= 0.72

Decision

split	Entropy	I.G
Perf	0.95	0.05
Class	0.72	0.28

(4) Reduction in variance-

$$\text{Mean} = \frac{2+6+7+4+7+9+12}{7}$$

$$= \frac{47}{7} = 6.7$$

Mean: $\frac{6}{6} = 1$

$(\text{Var})_B = 0$

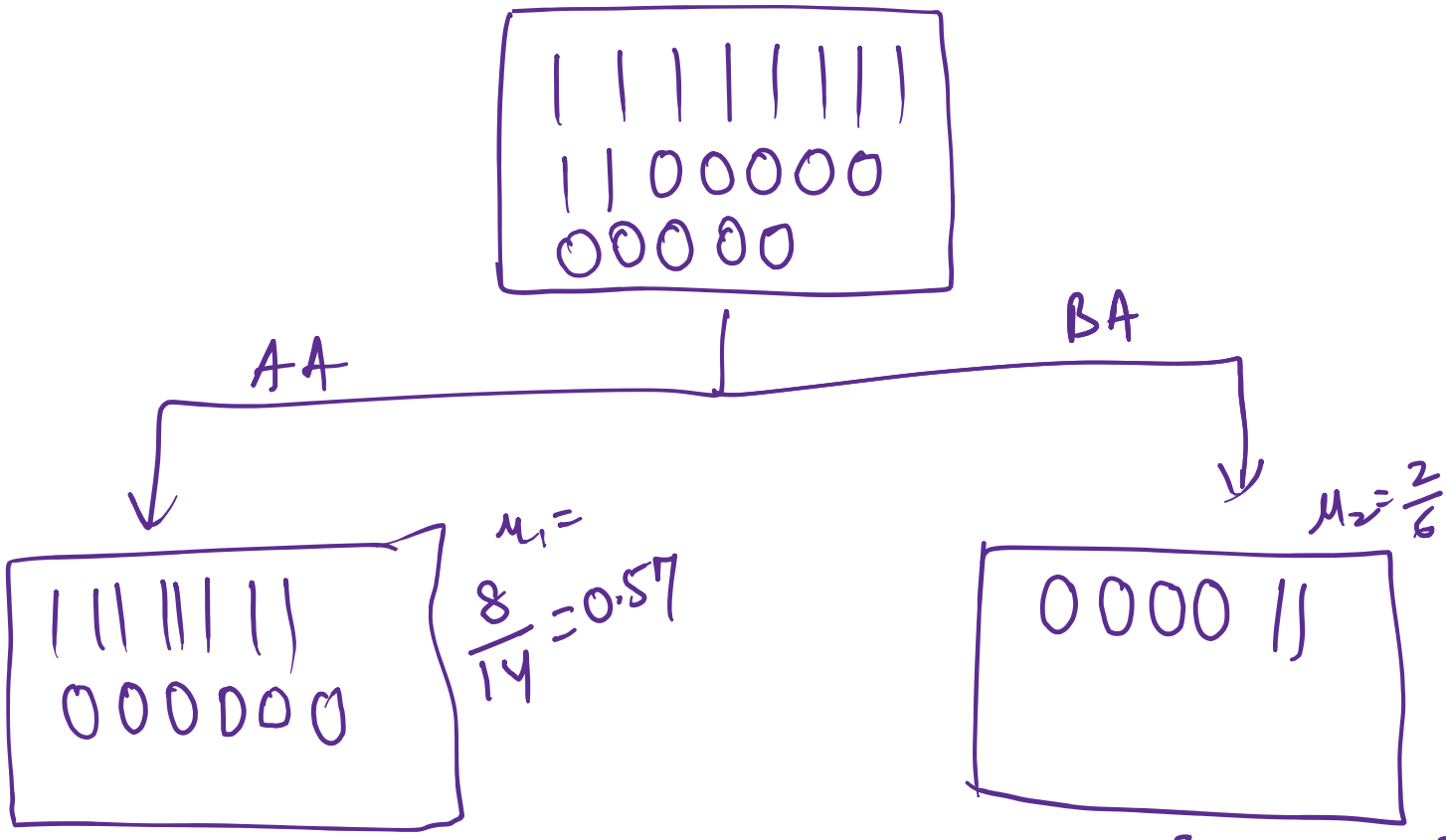
$$(\text{Var})_A = \frac{\Sigma (X-\mu)^2}{n}$$

$$(2-6)^2 + (6-6)^2$$

$$\frac{67}{7} = 9.$$

Lesser the variance → more Pure node
→ Homogeneous

Split 1 : Performance



① $V(AA) = \frac{\Sigma(x-\mu)^2}{n} = \frac{8(1-0.57)^2}{14} + \frac{6(0-0.57)^2}{14}$

$= 0.24$

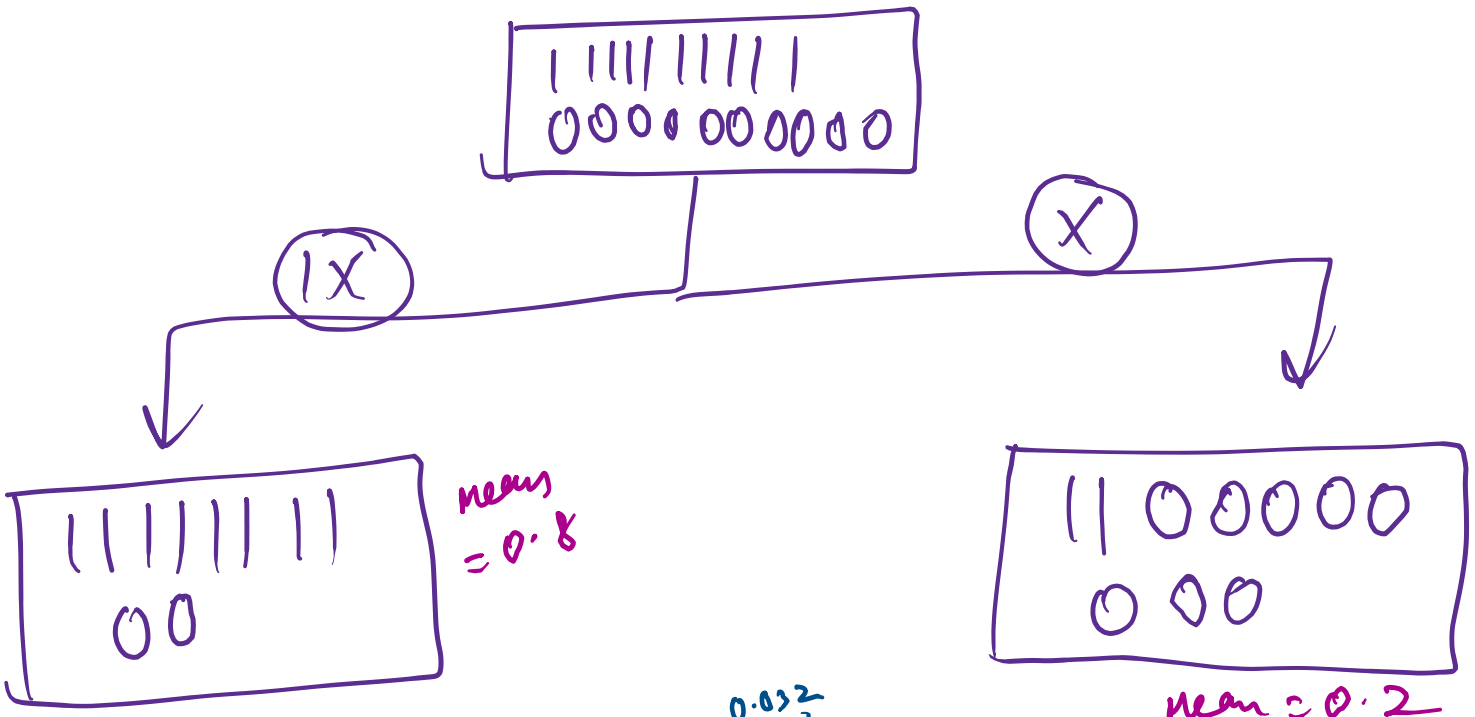
② $V(BA) = \frac{\Sigma(x-\mu)^2}{n} = \frac{2(1-0.33)^2}{6} + \frac{4(0-0.33)^2}{6}$

$= 0.22$

③ Weighted Var. $= \frac{14}{20} \times (0.24) + \frac{6}{20} \times (0.22)$

$= 0.23$

Split 2: Split by Class



0.032
0.128

① $\text{Var}(1X) = \frac{8(1-0.8)^2}{10} + \frac{2(0-0.8)^2}{10}$ $= 0.16$

② $\text{Var}(X) = \frac{2(1-0.2)^2}{10} + \frac{8(0-0.2)^2}{10}$

$= 0.16$

(3) weighted var. $= \frac{10}{20} \times (0.16) + \frac{10}{20} \times (0.16)$

$= 0.16$

decision

split	w. var.
Pet	0.24
Clan	0.16

EOF>>>>>>>>>>>>>>>>>>>>>>