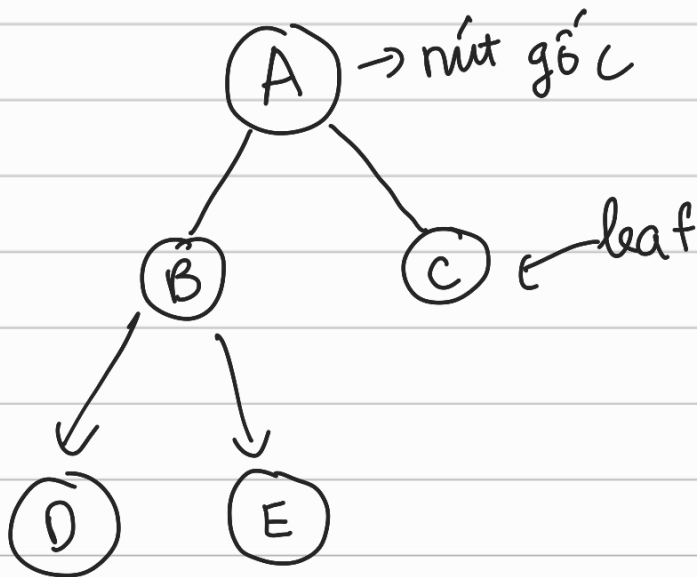
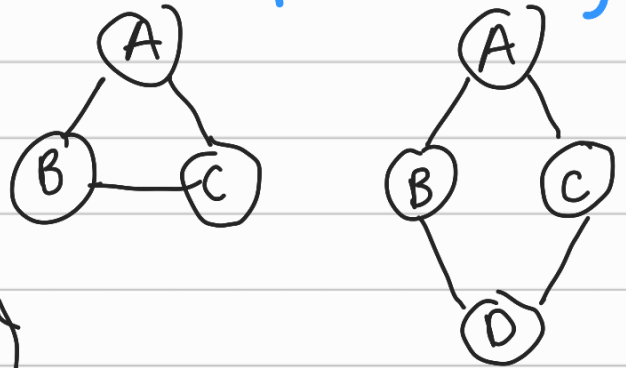


Decision Trees (most important).

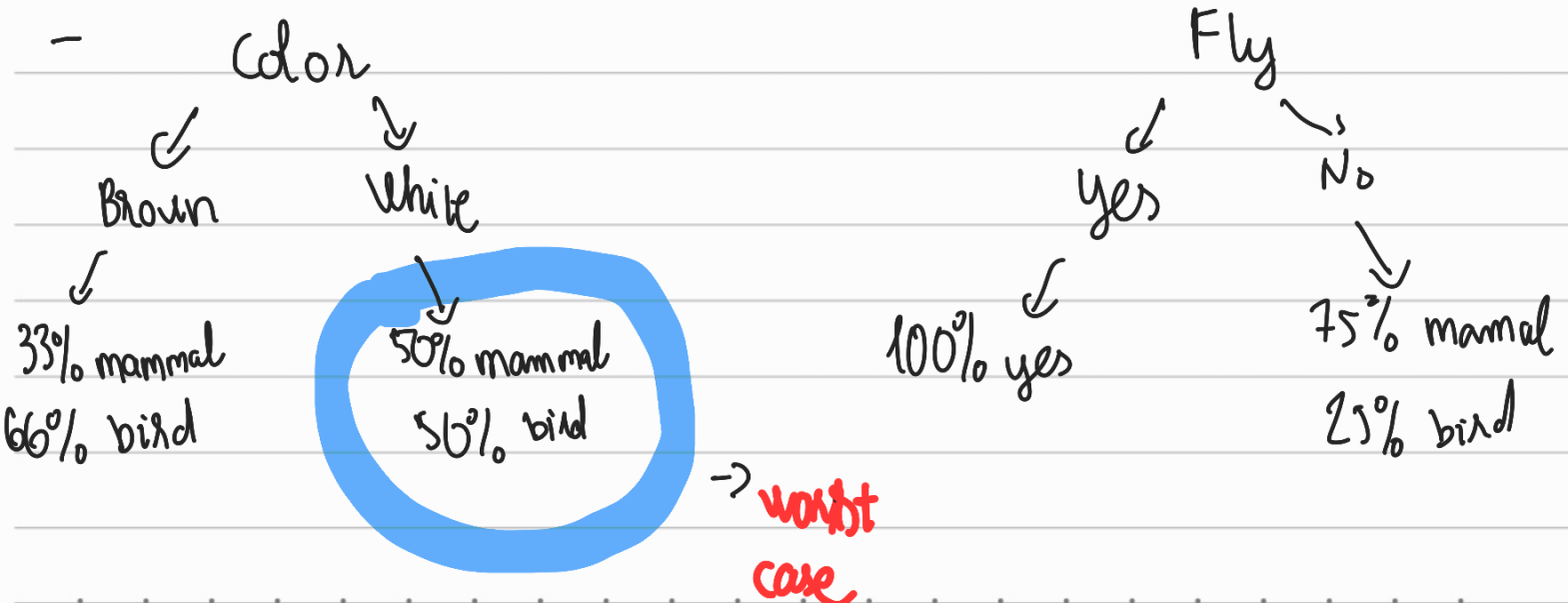
- Cây
 - root : no - parent (chỉ có 1 root)
 - node : qhệ cha - con
 - leaf : no child
- o có circle



đây o phải là cây:



- Ở DT, build 1 cái cây, các node o phải là để rẽ nhánh
- các node lá sẽ là gtrị dự đoán.



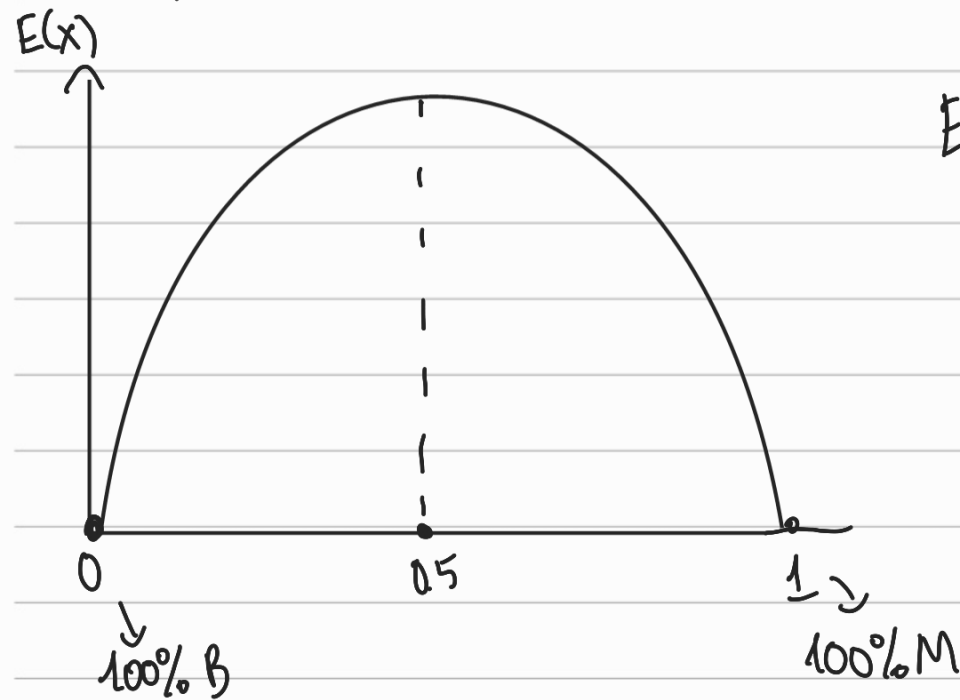
Hàm entropy: $E(X) = - \sum_{i=1}^k p_i \cdot \log_2 p_i$.

$$= - P(M) \cdot \log P(M) - P(B) \cdot \log P(B).$$

$$P(M) = x$$

$$P(B) = 1 - x$$

$$= -x \log x - (1-x) \log (1-x).$$



Entropy càng nhỏ
→ dự đoán càng tốt.

Độ giảm Entropy (càng lớn càng tốt)

$$\underbrace{E(P)}_{\text{entropy tốt nhất}} - \underbrace{E(c_1)}_{\text{con 1}} - \underbrace{E(c_2)}_{\text{con 2}}$$

| CT có vậ. |

$$\Rightarrow \text{gain}(X, a) = \text{entropy}(P) - \frac{|c_1|}{N} \cdot E(c_1) - \frac{|c_2|}{N} \cdot E(c_2)$$

$$N = |c_1| + |c_2|.$$

Attribute: Color = Brown, White
 $S = [3+, 4-]$.

$$\text{Entropy}(S) = -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) \\ = 0.985.$$

Entropy(Brown) $[1+, 2-]$.

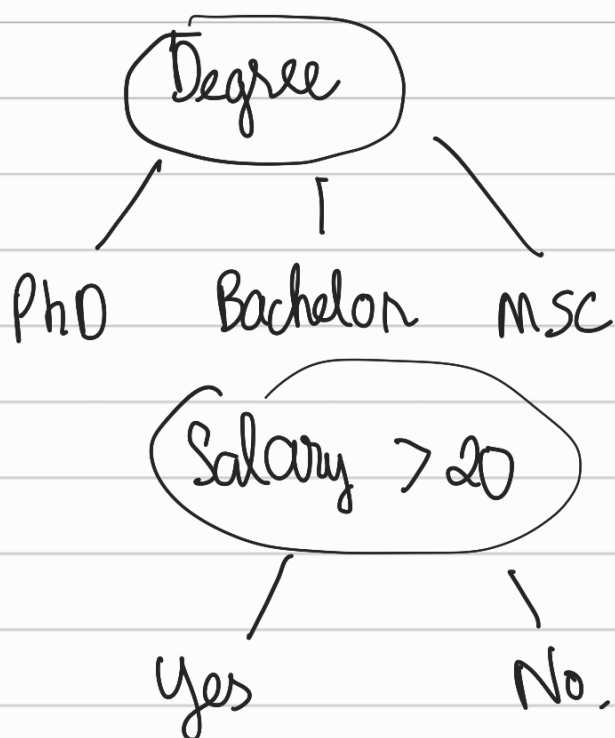
$$\text{Entropy}(\text{brown}) = -\frac{1}{3} \log_2\left(\frac{1}{3}\right) + \frac{2}{3} \log_2\left(\frac{2}{3}\right)$$

(3)

(4)

$$\text{Gini } gini(X) = 1 - \sum_{i=1}^k p_i^2$$

Salary	Degree	Loan
10	PHD	Yes.
20	Bachelor	No
15	MSC	Yes



/

/