

Q1

$$\begin{aligned}
 a) \quad d_1 &= \sqrt{(2-0)^2 + (3-0)^2 + (1-0)^2} = 3 \\
 d_2 &= \sqrt{(2-0)^2 + (0-0)^2 + (1-0)^2} = 2 \\
 d_3 &= \sqrt{(-1-0)^2 + (0-0)^2 + (1-0)^2} = \sqrt{2}
 \end{aligned}$$

$$d_3 = \sqrt{(0-0)^2 + (1-0)^2 + (3-0)^2} = \sqrt{10}$$

$$d_4 = \sqrt{(0-0)^2 + (1-0)^2 + (2-0)^2} = \sqrt{5}$$

$$d_6 = \sqrt{(1-0)^2 + (1-0)^2 + (1-0)^2} = \sqrt{3}$$

b) If  $K=1 \rightarrow$  Nearest Neighbor = Obs. 5  $\rightarrow$  prediction = Green

c) if  $K=3 \rightarrow 3$ -Nearest Neighbors =  $\{5, 6, 2\}$  <sup>Obs.</sup>  $\rightarrow Y(2) = \text{Red}$   
 for  $(0,0,0)$   $Y(5) = \text{Green}$   
 $Y(6) = \text{Red}$

prediction for  $(0,0,0)$  is Red  
(voting by its 3-Nearest Neighbors)

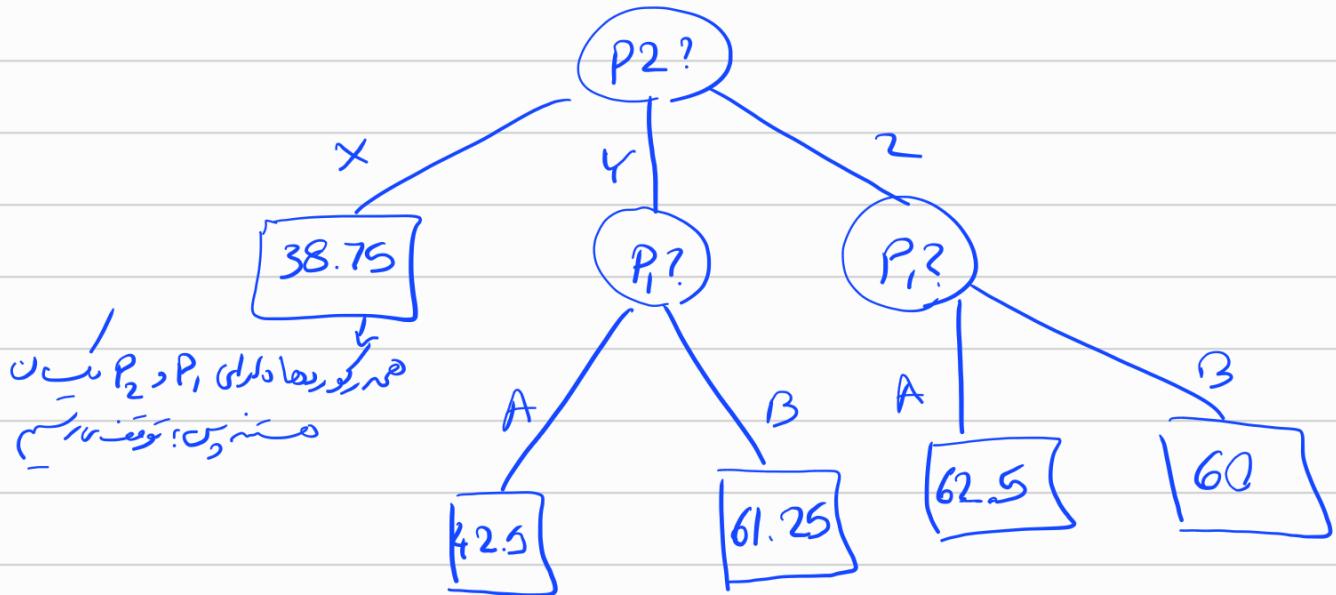
$$S_1: \text{Predictor 1} \quad \begin{matrix} A \\ B \end{matrix} \quad \hat{y}_e = \frac{30+40+60+45+50+65+35+40}{8} = 45.625$$

$$\hat{y}_{R_2} = \frac{50+55+70+60+80+50}{6} \approx 60.83$$

$$RSS(S_1 \mid \underline{\underline{y}}) = \sum_{i \in R_1} (y_i - \hat{y}_{R_1})^2 + \sum_{j \in R_2} (y_j - \hat{y}_{R_2})^2 \approx 1742.71$$

$$\begin{aligned}
 S_2, \text{ Predictor 2} & \quad X \quad \hat{y}_{R_1} = \frac{30+40+50+35}{4} = 38.75 \\
 & \quad Y \quad \hat{y}_{R_2} = \frac{50+55+45+60+80+40}{6} = 55 \\
 & \quad Z \quad \hat{y}_{R_3} = \frac{60+70+65+50}{4} = 61.25
 \end{aligned}$$

$$\text{RSS}(S_2) \stackrel{?}{=} \sum_{j=1}^3 \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 = 1437.5 \rightarrow \text{صحیح} \checkmark$$



## 2- Analyze

support:

4  $\sim 38.75$  target  $\leftarrow$   $X \rightarrow P_2$  ✓

2  $\sim 42.5$  target  $\leftarrow$   $P_1 = A \rightarrow P_2 = Y$  ✓

4  $\sim 61.25$  target  $\leftarrow$   $P_1 = B \rightarrow P_2 = Y$  ✓

2  $\sim 62.5$  target  $\leftarrow$   $P_1 = A \rightarrow P_2 = Z$  ✓

4  $\sim 60$  target  $\leftarrow$   $P_1 = B \rightarrow P_2 = Z$  ✓

$P_1$   $\sim$  sub-target  $(A, B, Y, Z)$   $P_2$   $\sim$   $(38.75, 55, 61.25, 60)$

part 2)

$> 51 \rightarrow \text{high}$

P 9

Predictor 1 (Categorical)	Predictor 2 (Categorical)	Target Variable (Numerical)	Target (Categorical)
A	X	30	L
B	Y	50	L
A	X	40	L
A	Z	60	H
B	Y	55	H
A	Y	45	L
B	Z	70	H
A	X	50	L
B	Y	60	H
A	Z	65	H
B	Y	80	H
A	X	35	L
B	Z	50	L
A	Y	40	L

part 3) 2) CART

$$\phi_{CS}(t) = 2P_L P_R \sum_{j=1}^2 \left| P(j|t_L) - P(j|t_R) \right|$$

$$\begin{array}{c}
 \text{S}_1: P \begin{array}{l} \diagup A \\ \diagdown B \end{array} \\
 P_L = \frac{8}{14} \quad \left\{ \begin{array}{l} L: 6 \\ H: 2 \end{array} \right. \quad P(L|t_L) = \frac{6}{8} \\
 P_R = \frac{6}{14} \quad \left\{ \begin{array}{l} L: 2 \\ H: 4 \end{array} \right. \quad P(H|t_R) = \frac{2}{6} \\
 P(L|t_L) = \frac{2}{6} \\
 P(H|t_R) = \frac{4}{4}
 \end{array}$$

$$\phi(S_1(t)) = 2 \times \frac{4}{7} \times \frac{3}{7} \left( \left| \frac{3}{4} - \frac{1}{3} \right| + \left| \frac{1}{4} - \frac{2}{3} \right| \right) = \frac{24}{49} \left( \frac{5}{12} + \frac{5}{12} \right) = \frac{20}{49}$$

$$\begin{array}{c}
 \text{S}_2 : q \quad \begin{array}{c} x \\ y, z \end{array} \\
 \hline
 P_L = \frac{4}{14} \quad \left\{ \begin{array}{l} L : 4 \\ H : 0 \end{array} \right. \quad \begin{array}{l} P(L|t_L) = 1 \\ P(H|t_R) = 0 \end{array} \\
 P_R = \frac{10}{14} \quad \left\{ \begin{array}{l} L : 4 \\ H : 6 \end{array} \right. \quad \begin{array}{l} P(L|t_L) = 0.4 \\ P(H|t_R) = 0.6 \end{array}
 \end{array}$$

$$\phi(S_2(t)) = 2 \cdot \frac{2}{7} \cdot \frac{5}{7} (11 - 0.4t + (0.4t)) = \frac{20}{49} (11.2) = \frac{24}{49}$$

$$S_3: q$$

$P_L = \frac{6}{14} \quad \begin{cases} L : 3 \\ H : 3 \end{cases} \quad P(L|t_L) = \frac{1}{2}$   
 $P_R = \frac{8}{14} \quad \begin{cases} L : 5 \\ H : 3 \end{cases} \quad P(H|t_R) = \frac{1}{2}$   
 $P(L|t_L) = \frac{5}{8}$   
 $P(H|t_R) = \frac{3}{8}$

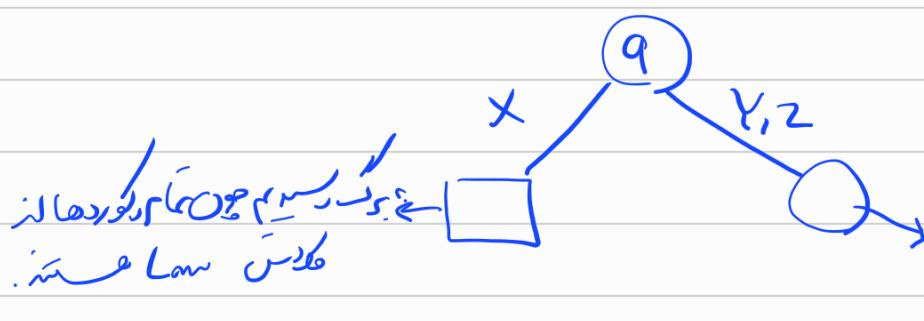
$$\Phi(S_3|t) = 2 \times \frac{3}{7} \times \frac{4}{7} \left( \left| \frac{1}{2} - \frac{5}{8} \right| + \left| \frac{1}{2} - \frac{3}{8} \right| \right) = \frac{24}{49} \left( \frac{1}{8} + \frac{1}{8} \right) = \frac{6}{49}$$

$$S_4: q$$

$P_L = \frac{4}{14} \quad \begin{cases} L : 1 \\ H : 3 \end{cases} \quad P(L|t_L) = \frac{1}{4}$   
 $P_R = \frac{10}{14} \quad \begin{cases} L : 7 \\ H : 3 \end{cases} \quad P(H|t_R) = \frac{3}{4}$   
 $P(L|t_L) = \frac{7}{10}$   
 $P(H|t_R) = \frac{3}{10}$

$$\Phi(S_4|t) = 2 \times \frac{2}{7} \times \frac{5}{7} \left( \left| \frac{1}{4} - \frac{7}{10} \right| + \left| \frac{3}{4} - \frac{3}{10} \right| \right) = \frac{20}{49} \left( \frac{9}{20} + \frac{9}{20} \right) = \frac{18}{49}$$

مهماتي في قوى ملحوظة  $S_2$  (ل)  $\Phi$



$$S_1: p$$

$P_L = \frac{4}{10} \quad \begin{cases} L : 2 \\ H : 2 \end{cases} \quad P(L|t_L) = \frac{1}{2}$   
 $P_R = \frac{6}{10} \quad \begin{cases} L : 2 \\ H : 4 \end{cases} \quad P(H|t_R) = \frac{1}{2}$   
 $P(L|t_L) = \frac{1}{3}$   
 $P(H|t_R) = \frac{3}{3}$

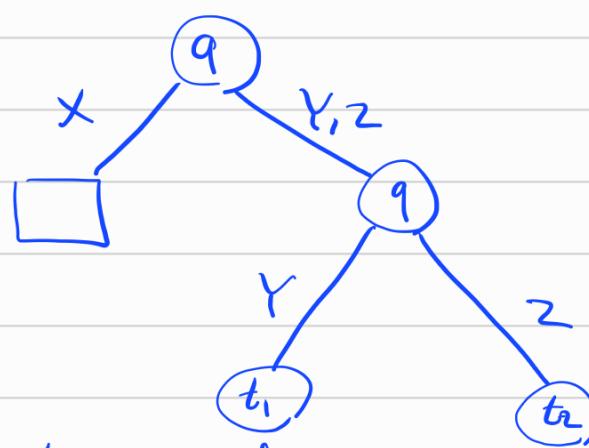
$$\Phi(S_1|t) = 2 \times \frac{4}{10} \times \frac{6}{10} \left( \left| \frac{1}{2} - \frac{1}{3} \right| + \left| \frac{1}{2} - \frac{2}{3} \right| \right) = \frac{48}{100} \left( \frac{1}{3} \right) = 0.16$$

$$S_2: q$$

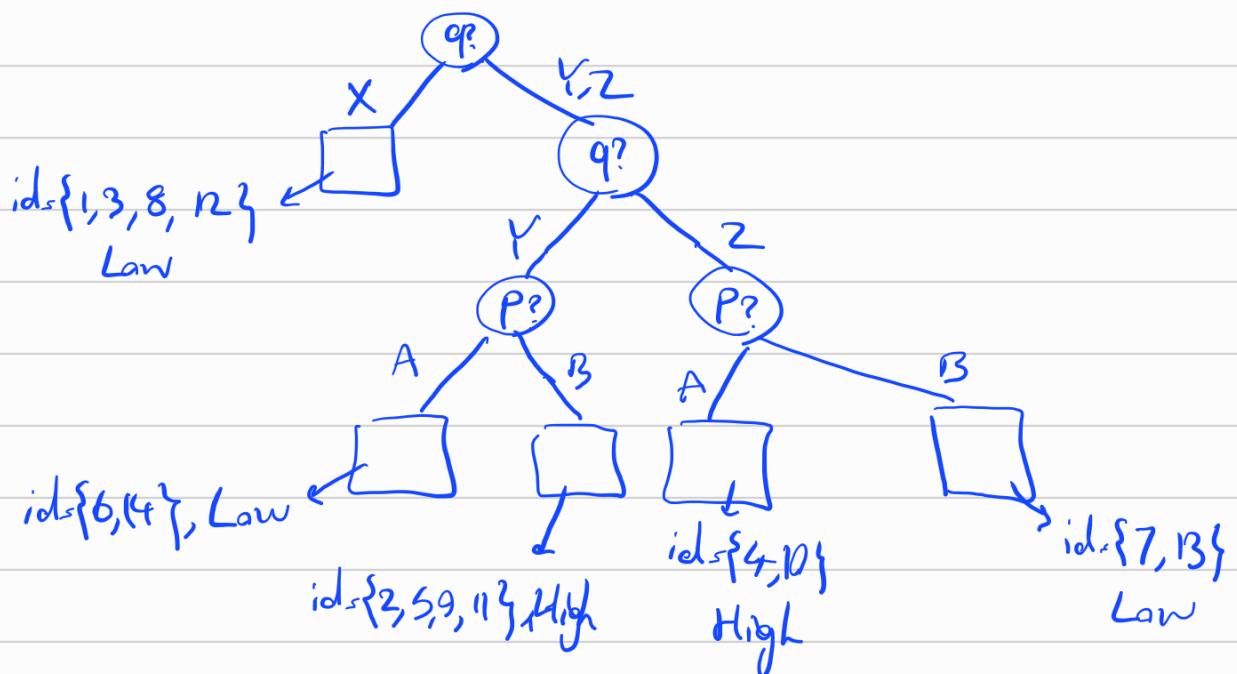
$P_L = \frac{6}{10} \quad \begin{cases} L : 3 \\ H : 3 \end{cases} \quad P(L|t_L) = \frac{1}{2}$   
 $P_R = \frac{4}{10} \quad \begin{cases} L : 1 \\ H : 3 \end{cases} \quad P(H|t_R) = \frac{1}{2}$   
 $P(L|t_L) = \frac{1}{4}$   
 $P(H|t_R) = \frac{3}{4}$

$$\phi(S_2(t)) = 2 \times \frac{6}{10} \times \frac{4}{10} (|1\frac{1}{2} - \frac{1}{4}| + |1\frac{1}{2} - \frac{3}{4}|) = \frac{48}{100} (1\frac{1}{2}) = \frac{24}{100}$$

نحوه سیمی  $S_2$  ۰٪



لذین دو سه صفحه در سه پیکسل های موجود در  $t_1$  و  $t_2$  ترکیب شوند  $P$  کوچک شوند  $q=2, q=2$



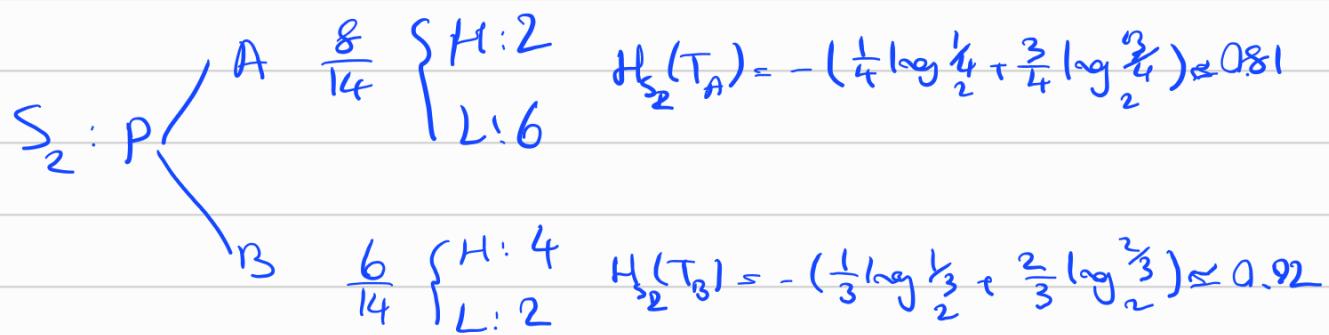
C 4.5 Method)

$$\begin{aligned}
 & X \quad \frac{4}{14} \quad \begin{cases} H: 0 \\ L: 4 \end{cases} \quad H_S(T_X) = -(1 \log_2 1 + 0 \log_2 0) = 0 \\
 \therefore q: & Y \quad \frac{6}{14} \quad \begin{cases} H: 3 \\ L: 3 \end{cases} \quad H_S(T_Y) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1 \\
 & Z \quad \frac{4}{14} \quad \begin{cases} H: 3 \\ L: 1 \end{cases} \quad H_S(T_Z) = -\left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4}\right) = 0.81
 \end{aligned}$$

$$H(T) = -\left(\frac{8}{14} \log_2 \frac{8}{14} + \frac{6}{14} \log_2 \frac{6}{14}\right) \leq 0.985$$

$$H_{S_1}(T) = \sum_{i=1}^3 p_i H_{S_1}(T_i) = \frac{4}{14} \times 0 + \frac{6}{14} \times 1 + \frac{4}{14} \times 0.81 = 0.66$$

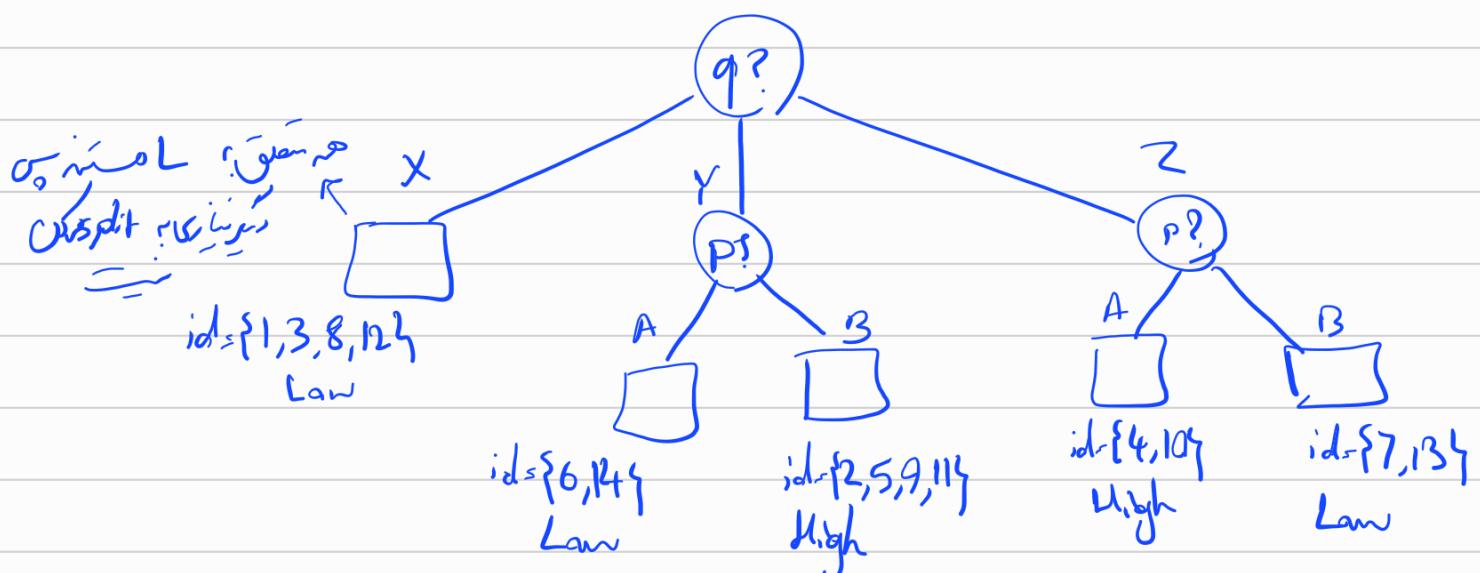
$$g_{\text{avH}}(S_1) = H(T) - H_{S_1}(T) = 0.985 - 0.66 = 0.325$$



$$H_2(T) = \sum_{i=1}^2 P_i H_2(T_i) = \frac{8}{14} \times 0.81 + \frac{6}{14} \times 0.92 = 0.85$$

$$g_{\text{air}}(S_2) = H(T) - H(T) \frac{0.985 - 0.85}{S_2} = 0.135$$

هـ اسپیت اول نسب تریت چون بیان ماهن انتقای بینه دارد سینه بیان سب اصلاح بالتریت.



## Part 4)

Miss Classification Error for C4.5:  $\frac{2}{14} = \frac{1}{7}$  id. §2, 73

~ ~ ~ for CART:  $\frac{2}{14}, \frac{1}{7}, \text{AdS}\{2,7\}$

Questions:

1) خصی مددیون تولید کنندگان C4.5 و CART چه میان اینها تفاوتی وجود ندارد (۱) و از اینها کدامیک برای کاربرد در میان این دو مدل انتخاب شود (۱)

2) Miss classification Error چه میان این دو مدل انتخاب شود (۱) و از اینها کدامیک برای کاربرد در میان این دو مدل انتخاب شود (۱)

3) در میان این دو مدل کدامیک برای کاربرد در میان این دو مدل انتخاب شود (۱) و از اینها کدامیک برای کاربرد در میان این دو مدل انتخاب شود (۱)

4) در میان این دو مدل کدامیک برای کاربرد در میان این دو مدل انتخاب شود (۱) و از اینها کدامیک برای کاربرد در میان این دو مدل انتخاب شود (۱)

5) Feature Selection - 1

6) Pruning - 2

7) Cross Validation - 3