Pujitha Kallu
W1653660

Problem 1:

a) 3 [dimensionality of dataset]

b) [8.2 2.5 1.7] (x^5)

c) 6.9 (x_1^3)

d) 0.6 (x_3^1)

e) blue (y^4)

Problem 2:

a) 1, 2 [values of i]

b) 3 [value of d]

c) 3x1 [dimension of μ_i]

d) 3 x 3 [dimension of Σ_i]

e) 1/2 [value for $P(\omega_i)$] for equal piror probabilities probabilities.

Problem 3:

A) d Logistic regression

B) e Elastic regression

C) c High generalization error

D) c Logistic regression

E) a stop training when validation error reaches minimum

F) c 2

G) c softmax Regression.

H) C It is susceptible to outliers

Problem 4:

Sol: $TP = 5$ [~~total~~ True positive]

$FN = 4$ [False Negatives].

$FP = 3$ [False positives]

$$\text{Precision} = \frac{TP}{(TP+FP)} = \frac{5}{(5+3)} = \frac{5}{8} = 62.5\%$$

$$\text{Recall} = \frac{TP}{(TP+FN)} = \frac{5}{(5+4)} = \frac{5}{9} = 55.6\%$$

$\therefore$ $\boxed{\begin{array}{l}\text{Precision} = 62.5\% \\ \text{Recall} = 55.6\%\end{array}}$

Problem 5:

a) Given $w = [1, -2, 3, -4]$

$x = [10, 20, 30, 40]$

Predicted value $\hat{y} = w \cdot x$

$\hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4$

$\hat{y} = 1 \times 10 + (-2)(20) + (3)(30) + (-4)(40)$

$\hat{y} = 10 - 40 + 90 - 160$

$\boxed{\hat{y} = -100}$

b) Ridge Regression :

$$J(\omega) = MSE(\omega) + \frac{1}{2} \alpha \sum_{i=1}^{n} w_i^2$$

Penality term of Ridge Regression when parameter set to 0.6 is :

$$= \frac{1}{2} \times 0.6 \left(1^2 + (-2)^2 + (3)^2 + (-4)^2\right)$$

$$= 0.3 (1 + 4 + 9 + 16)$$

$$= 0.3(30) = 9$$

∴ Penality term of Ridge Regression is $\underline{\underline{9}}$

Problem 6 :

Given data set $X = [50, 60, 70, 80]$

$Y = [c, a, c, b]$

	X	Y
55	50	c
65	60	a
75	70	c
	80	b

X	Y=a	Y=b	Y=c
50			1
60	1		
70			1
80		1	

Maximum depth = 1 [root node & its 2 childen].

$$Gini = 1 - \sum_{i=1}^{m} P_i^2$$

$$= 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{1}{4}\right)^2 - \left(\frac{2}{4}\right)^2$$

$$Gini = 1 - 0.0625 - 0.0625 - 0.25 = 1 - 0.375$$

$$Gini = 0.625$$

Scanned with CamScanner

a) Threshold = 55

$\boxed{x < 55}$

Y

a	b	c
0	0	1

a	b	c
$\cancel{0}$	1	1

Gini Impurity for Left Node =

$$\text{Gini}_{Left} = 1 - (\text{Prob of } a)^2 - (\text{Prob of } b)^2 - (\text{Prob of } c)^2$$

$$\text{Gini left} = 1 - \left(\frac{1}{1}\right)^2 = 0$$

$$\underline{\text{Gini}_{Left} = 0}$$

Gini Impurity for Right Node =

$$\text{Gini}_{Right} = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{1}{3}\right)^2 - \left(\frac{1}{3}\right)^2$$

$$= 1 - 0.1111 - 0.1111 - 0.1111$$

$$= 1 - 0.4444$$

$$\underline{\text{Gini}_{Right} = 0.6667}$$

Minimum CART cost function [Combined Gini Impurity]

$$J(k, T_L) = \frac{m_{left}}{m} \text{Gini}_{left} + \frac{m_{right}}{m} \text{Gini}_{right}$$

$$= \left(\frac{1}{4}\right)(0) + \left(\frac{3}{4}\right)(0.6667)$$

$$\boxed{J(k, T_K) = 0.500025}$$

for Gini_{right} 0.667 => $J(k, T_K) = 0.500025$

Threshold = 65

$\boxed{X < 65}$

y   y

a	b	c
1	0	ϕ

a	b	c
0	1	1

$$\text{Gini Left} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - 0.25 - 0.25 = 1 - 0.5$$

$$\text{Gini Left} = 0.5$$

$$\text{Gini Right} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - 0.25 - 0.25 = 1 - 0.5$$

$$\text{Gini Left} = 0.5$$

Minimum CART function:

$$J(k, T_k) = \left(\frac{2}{4}\right)(0.5) + \left(\frac{2}{4}\right)(0.5)$$

$$\boxed{J(k, T_k) = 0.5}$$

Threshold = 75

$\boxed{X < 75}$

a	b	c
1	0	2

a	b	c
0	0	1

$$\text{Gini left} = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$$

$$= 1 - 1/9 - 4/9 = 0.444$$

$$\boxed{\text{Gini Left} = 0.444}$$

$$Gini_{right} = 1-\left(\frac{1}{1}\right)^2 = 0$$

$$Gini_{right} = 0$$

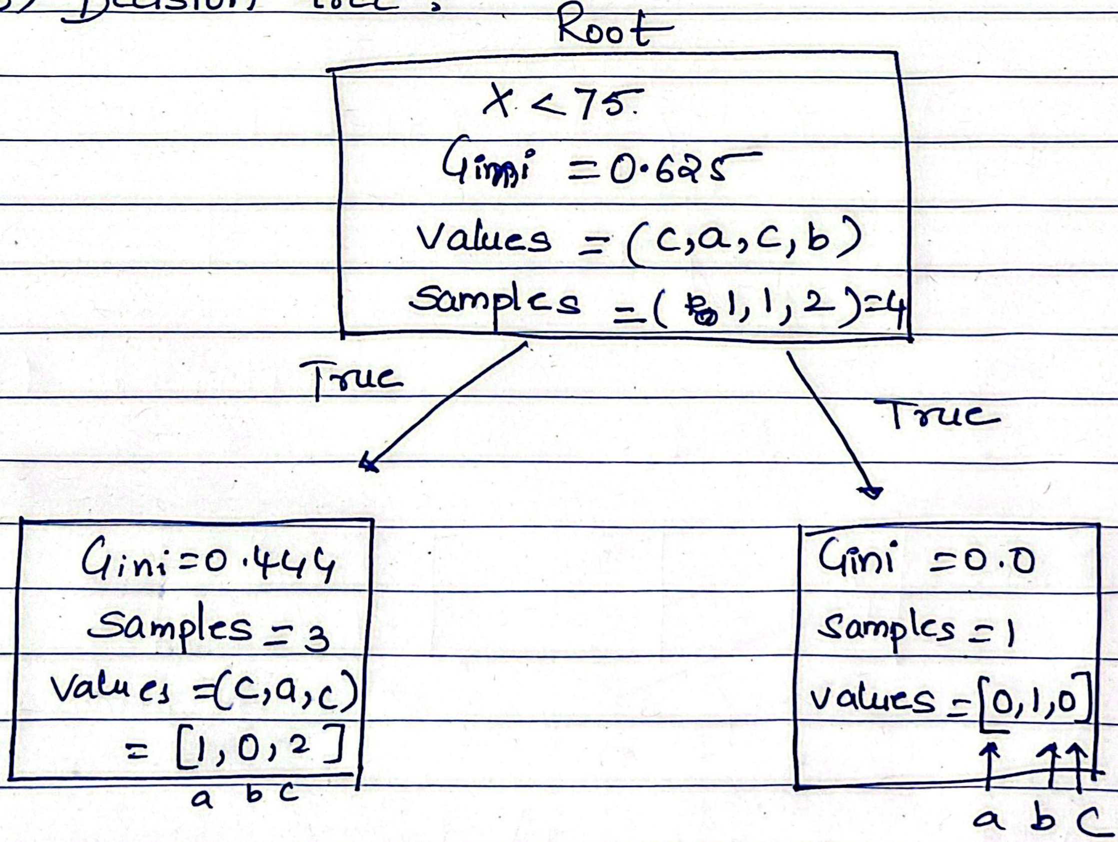
Minimum CART function

$$J(k, T_k) = \left(\frac{3}{4}\right)(0.444) + \left(\frac{1}{4}\right) 0$$

$$\boxed{J(k, T_k) = 0.333}$$

From the above threshold we get
Minimum threshold for 75
Cost value = 0.333.

b) Decision tree :



Gini squo score of root $= 1 - \sum_{i=1}^{m} p_i^2$

$$= 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{1}{4}\right)^2 - \left(\frac{2}{4}\right)^2$$

$$= 1 - \left(\frac{1}{16}\right) - \left(\frac{1}{16}\right) - \left(\frac{1}{4}\right)$$

Gini score of root $= 0.625$