

$$Q1: Y = X_1 + 2X_2 + 3X_3 + 4X_4$$

$$X_1, X_2, X_3, X_4 \in [-2, 2]$$

proposed $\hat{Y} = 10X_1$

$$MSE = ?$$

let us assume that

$$\left. \begin{array}{l} X_1 = -1.5 \\ X_2 = -1 \\ X_3 = 0 \\ X_4 = 1 \end{array} \right\} \text{real entries in range } (-2, 2)$$

$$Y = -1.5 - 2 + 4 = 0.5$$

$$\hat{Y} = 10(-1.5) = -15$$

$$MSE = \frac{1}{n} (\hat{Y} - Y)^2$$

when $n=1$

$$= \frac{1}{1} (15.5)^2 = 240.25$$

deriving expression for MSE:

$$MSE = \frac{1}{n} (\hat{Y} - Y)^2 = (10X_1 - X_1 - 2X_2 - 3X_3 - 4X_4)^2$$

$$= (9X_1 - 2X_2 - 3X_3 - 4X_4)^2$$

18-8-3-2

$$0 \leq MSE \leq (-18 - 9(2))^2 \rightarrow 1296$$

for all X_2, X_3, X_4

and $X_1 = -2$

for all $X_2, X_3, X_4 = 2$,
and $X_1 = -2$.

$$0 \leq MSE \leq 1296$$

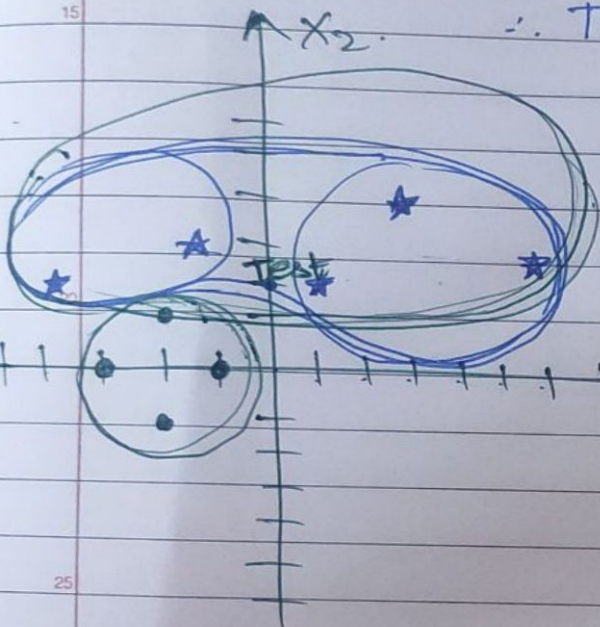
$$= |X_1 - X_2|$$

using Manhattan distance formula

Q2. $k=1$

ID	X_1	X_2	Y	1	2	3	4	5	6	7	8	9	Test
1	-1	1	A	1	1	3	7	2	9	3	4	3	2
2	-1	0	A	2		2	8	3	10	2	5	4	3
3	-2	-1	A	3			10	5	12	2	5	6	5
4	3	4	B	4				5					5
5	-1	3	B	5									2
6	6	3	B	6									7
7	-3	0	A	7									5
8	-4	2	B	8									4
9	1	2	B	9									1
Test:	0	2	?										

for $k=1$, Test data is nearest to 9th sample
 \therefore Test label = (B)



• A.
 ★ B.

made cluster of 2.
 Test point comes
 in cluster of (B).

Q3. KNN, Decision Stump.

ID	X ₁	X ₂	Y	Manhattan Distance (9)	(10)
1	-1	4	A	7	5
2	5	12	B	7	9
3	-5	-2	A	17	15
4	-3	-5	A	18	16
5	6	10	B	6	8
6	2	6	A	(2)	(2)
7	-5	8	B	7	11
8	2	11	B	3	7
9	2	8	? → A	(B)	
10	3	5	? → A	(B)	

KNN Decision Stump.

taking split of $X_1 > 0$.

	$X_1 > 0$	$X_1 \leq 0$	No split
Y = A	1	3	4
Y = B	3	1	4
entropy	0.811	0.811	(1) original entropy

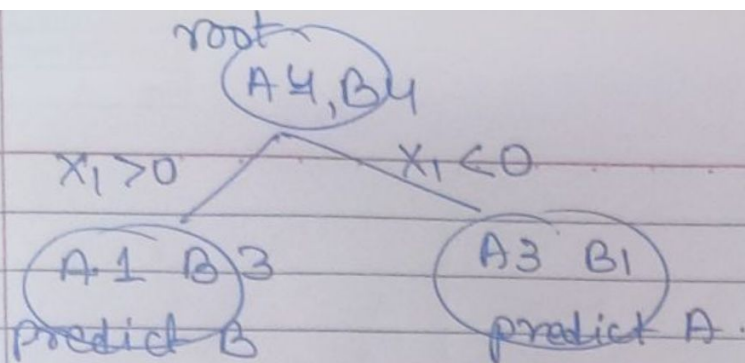
$\rightarrow -P_i \log_2 P_i$
 $- [Y_4 \log_2 Y_4 + 3/4 \log_2 3/4]$
 $- [Y_2 \log_2 Y_2 + Y_2 \log_2 Y_2]$

original entropy.

Information gain = (1) - split 1 * entropy split 1 - split 2 * entropy split 2.

$= 1 - Y_2(0.811) - Y_2(0.811)$

$= 1 - 0.811 = [0.189]$



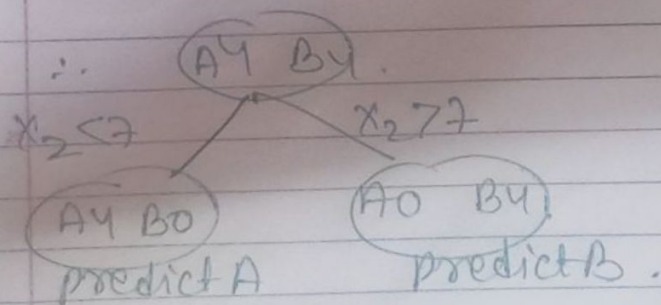
for 9th sample, $X_1 = 2, X_2 = 8$
 $\hat{Y} = \textcircled{B}$

for 10th sample, $X_1 = 3, X_2 = 5$
 $\hat{Y} = \textcircled{B}$

taking split of
 $X_2 > 7$.

	$X_2 \leq 7$	$X_2 > 7$	No split
$Y = A$	4	0	4
$Y = B$	0	4	4
entropy	0	0	1
	↑	↑	← $-\left[\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}\right]$
	← $-\left[\frac{4}{8} \log \frac{4}{8} + \frac{4}{8} \log \frac{4}{8}\right]$		

Information gain = $1 - \frac{1}{2}(0) - \frac{1}{2}(0) = \boxed{1}$



for 9th sample $\Rightarrow \hat{Y} = \textcircled{B}$
 for 10th sample $\Rightarrow \hat{Y} = \textcircled{A}$

∴ By changing split the information gain changes and so does the predicted value