Q1] $y = 1 \quad \therefore \quad [n]_{c_1} = 53 \quad T = 200$

$y = 2 \quad : \quad [n]_{c_2} = 59$

$y = 3 \quad : \quad [n]_{c_3} = 88$

Class entropy $= -\frac{53}{200} \log\left(\frac{53}{200}\right) + \frac{59}{200} \log\left(\frac{200}{59}\right)$

$+ \frac{88}{200} \log\left(\frac{200}{88}\right)$

$= 1.55$

Information gain = CE − total entropy of split

For x_1:	$y=1$	$y=2$	$y=3$	I_g
1	31	40	34	1.53
2	22	19	54	1.42

Entropy of x_1: $\frac{\Sigma [n]_{ci}}{\Sigma [n]} \log_2([n]_{c})$

$: \frac{105}{200} \times 1.53 + \frac{95}{200} \times 1.42 = 1.47$

For x_2:	$y=1$	$y=2$	$y=3$	I_g
1	25	26	45	1.53
2	28	33	43	1.56

Entropy of x_2: $\frac{96}{200} \times 1.53 + \frac{104}{200} \times 1.56$

$= 1.54$

for x_3	$y=1$	$y=2$	$y=3$	I_g
A	35	14	1	0.99
B	15	25	10	1.48
C	1	15	34	1.01
D	2	5	43	0.70

entropy of x_3 : $\frac{1}{4} \times 0.99 + \frac{1}{4} \times 1.48 + \frac{1}{8} \times 1.01$

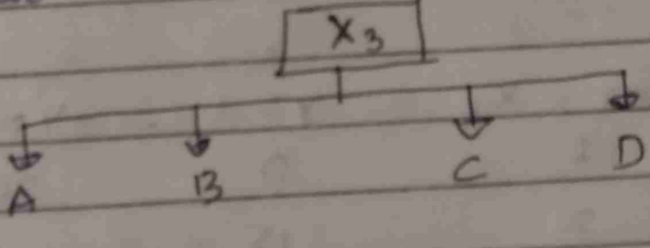
$+ \frac{1}{8} \times 0.7 = 1.04$

Gain of $x_1 = 1.55 - 1.47 = 0.08$

$x_2 = 1.55 - 1.54 = 0.01$

$x_3 = 1.55 - 1.04 = 0.51$

Root node : x_3



In A : Class entropy = 0.99

For x_1	$y=1$	$y=2$	$y=3$	I_g
1	30	0	0	0
2	5	14	1	1.08

entropy of x_1 in A = $\frac{20}{50} \times 1.08 = 0.432$

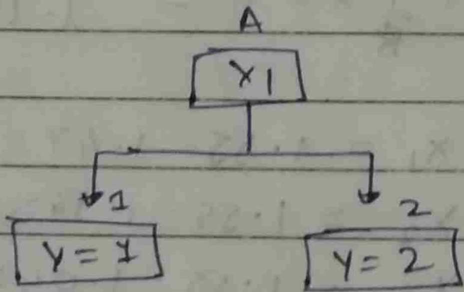
Gain of x_1 in A = 0.56

for x_2	$y=1$	$y=2$	$y=3$	I_g
1	18	5	0	0.76
2	17	9	1	1.32

Entropy of x_2 in A $= \frac{23}{50} \times 0.76 + \frac{27}{50} \times 1.32$

$= 1.06$

Gain of x_2 in A $= 0.99 - 1.06 = -0.07$



<u>In B</u>

for x_1	$y=1$	$y=2$	$y=3$	I_g
1	0	20	5	0.72
2	15	5	5	1.37

entropy of x_1 in B $= 1.045$

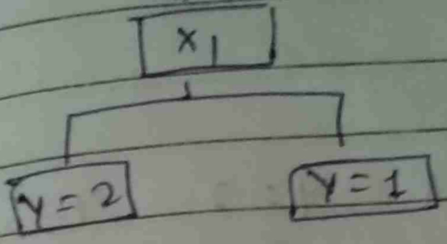
gain $= 1.48 - 1.045 = 0.435$

for x_2	$y=1$	$y=2$	$y=3$	I_g
1	7	13	5	1.47
2	8	12	5	1.5

entropy of x_2 in B : 1.485

gain $= -0.005$

B



In C:

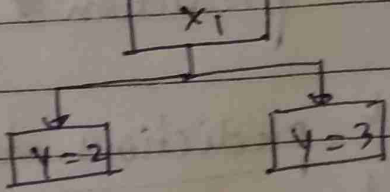
for x_1	y=1	y=2	y=3	Ig
1	0	15	0	0
2	1	0	34	0.18

entropy of x_1 in C = 0.126
Gain = 0.884

for x_2	y=1	y=2	y=3	Ig
1	0	6	20	0.78
2	1	9	14	1.18

entropy of x_2 in C = 0.97
Gain = 0.04

C



In D

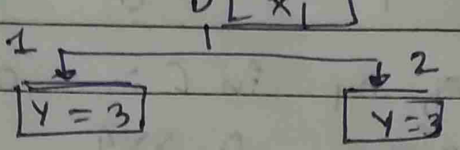
for x_1	y=1	y=2	y=3	Ig
1	1	5	9	0.77
2	1	0	14	0.35

entropy = 0.64 gain = 0.66

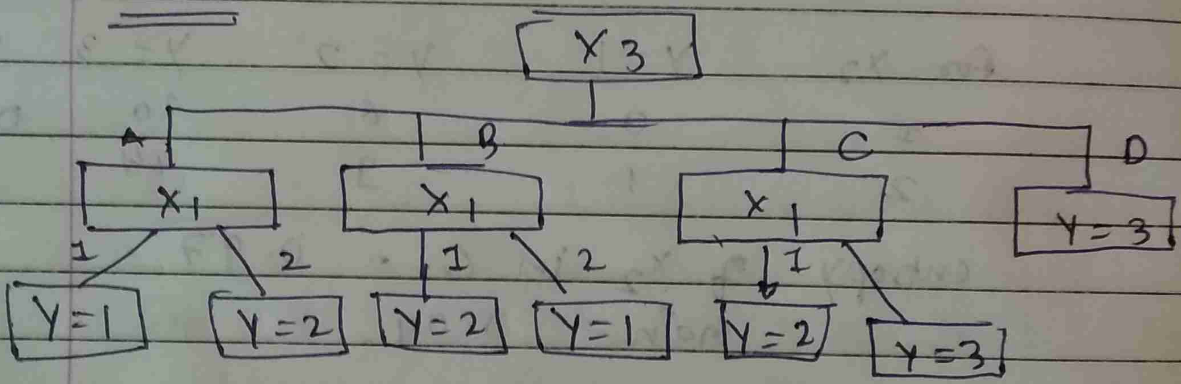
Date
Page

for x_2	$Y=1$	$Y=2$	$Y=3$	Ig
1	0	2	20	0.44
2	2	3	23	0.85

entropy = 0.67 Gain = 0.03



Tree:



Total correct prediction ⇒ $\frac{171}{200}$ accuracy

$$\Rightarrow \frac{\text{No. of correct predictions at each node}}{\text{Total data points}}$$

$$\Rightarrow \frac{171}{200}.$$

2. For finding the missing values:

1) Missing value ① $\Rightarrow$ when $x_1 = 2$, $x_2 = 1$, $x_3 = A$
Calculating individual probabilities for each x_i given $y = 1$

$$P(x_1 = 2 / y=1) = \frac{12}{43}$$

$$P(x_2 = 1 / y=1) = \frac{15}{43}$$

$$P(x_3 = A / y=1) = \frac{32}{43}$$

Using Naive Assumption

$$P(X / y=1) = P(x_1 = 2 / y=1) \cdot P(x_2 = 1 / y=1) \cdot P(x_3 = A / y=1)$$

$$= \frac{12 \times 15 \times 32}{(43)^3} = 0.0724$$

$$P(Y=1 / X) = \frac{P(X / y=1) \cdot P(y=1)}{P(X)}$$

$$= \frac{0.0724 \times 43/170}{\frac{71}{170} \times \frac{66}{170} \times \frac{42}{170}} = 0.4571$$

Similarly

$$P(X / y=2) = \frac{44}{3975} \quad P(Y=2 / X) = \underline{\underline{0.086}}$$

$$P(X/_{Y=3}) = \frac{48}{82} \times \frac{39}{82} \times \frac{1}{82} = \frac{234}{68921}$$

$$P(Y=3/_X) = 0.041$$

$\therefore$ Label prediction for $(x_1, x_2, x_3) \equiv (2, 1, A)$ is $\boxed{Y = 1}$ using mode and corresponding confidence is 0.457

II) Missing Value ② $\Rightarrow$ $(2, 1, B)$

$$P(x_1 = 2/_{Y=1}) = \frac{12}{43}$$

$$P(x_2 = 1/_{Y=1}) = \frac{15}{43}$$

$$P(x_3 = B/_{Y=1}) = \frac{8}{43}$$

Using Naive Bayes :

$$P(X/_{Y=1}) = \frac{12 \times 15 \times 8}{(43)^2} = 0.018$$

$$\therefore P(Y=1/_X) = \frac{0.018 \times 43/_{170}}{\frac{71 \times 66 \times 39}{(170)^3}} = 0.122$$

$$P(X/_{Y=2}) = \frac{11 \times 12 \times 22}{(45)^2} = 0.0318$$

$$P(Y=2/_X) = \underline{\underline{0.226}}$$

$$P(X/y=3) = \frac{48 \times 39 \times 9}{(82)^3} = 0.0305$$

$$P(y=3/x) = \frac{0.0305 \times 82/170}{\frac{71 \times 66 \times 39}{(170)^3}} = \underline{\underline{0.395}}$$

Hence for (2, 1, B) the label pred^n is $y=3$ with confidense 0.395.

(III) Missing Value (3) : (1, 1, C)

$$P(X/y=1) = \frac{31 \times 15}{(43)^2} = \frac{465}{79507}$$

$$P(y=1/x) = \underline{\underline{0.025}}$$

$$P(X/y=2) = 0.0403$$

$$P(y=2/x) = \frac{0.0403 \times 45 (170)^2}{99 \times 66 \times 44} = \underline{\underline{0.182}}$$

$$P(X/y=3) = 0.0817$$

$$P(y=3/x) = \frac{0.0817 \times 82 (170)^2}{99 \times 66 \times 44} = \underline{\underline{0.673}}$$

$\therefore$ For (1, 1, C) predicted label $\boxed{y=3}$
confidense = 0.673

(IV) Missing Value ⇒ (2, 1, D)

$$P(X/Y=1) = \frac{12 \times 15 \times 2}{(43)^3} = 4.53 \times 10^{-3}$$

$$P(Y=1/X) = \frac{4.53 \times 10^{-3} \times 42 (170)^2}{71 \times 66 \times 45}$$

$$= \underline{0.026}$$

$$P(X/Y=2) = \frac{11 \times 12 \times 5}{(45)^3} = 7.24 \times 10^{-3}$$

$$\therefore P(Y=2/X) = \frac{7.24 \times 10^{-3} \times \frac{45}{170} (170)^3}{71 \times 66 \times 45} = \underline{\underline{0.045}}$$

$$P(X/Y=3) = 0.128$$

$$P(Y=3/X) = \frac{0.128 \times 82/170}{\frac{71 \times 66 \times 45}{(170)^3}} = \underline{\underline{0.887}}$$

∴ for (2, 1, D) predicted label is $Y = 3$ with confidense = 0.887.

4th) $P(Y = / x_1 x_2) = \frac{P(x_1 x_2 / y) \, P(y)}{P(x_1 x_2)}$

$= k \, P(x_1/y) \, P(x_2/y) \, P(y)$

Now class conditionals are

1) $P(x_1 / y=1) = \nabla(\mu, \sigma) \cdot \frac{k_1}{2}$

$\mu = 8.9$

$s = 6.76$

$P(x_1 / y=1) = \frac{k_1}{2} \nabla(8.9, 6.76)$

2) $P(x_2 / y=1) \Rightarrow$

$(\mu = 9.2 \quad s = 7.288)$

$\Rightarrow \frac{k_2}{2} \nabla(9.2, 7.288)$

3) $P(x_1 / y=2) \Rightarrow \; \mu = 14.4$

$s = 4.933$

$\Rightarrow \nabla(14.4, 4.933) \cdot \frac{k}{2}$

4) $P(x_2/y=2) = \sigma(9.9, 16.544) \frac{P}{2}$

$\boxed{K1 = P = 1}$ $\boxed{\begin{array}{l} \mu = 9.9 \\ s = 16.544 \end{array}}$

Now for least confidense
for x_1, x_2

$P(y=1/x_1 x_2) = P(y=2/x_1 x_2)$

For a pair of (x_1, x_2) this should satisfy.

$\sigma(9.2, 7.288)\,\sigma(8.9, 6.76)$

$= \sigma(9.9, 16.544)\,\sigma(14.4, 4.93)$

3. Ans → Candidates from plot $x_1 = 11$ ~~[illegible]~~

For $x_1 = 11$

x_1	$y = 1$	$y = 2$	entropy
$x_1 \leq 11$	8	1	0.503
$x_1 > 11$	2	9	0.634

$$\text{Total } E = \frac{9}{20} \times 0.503 + \frac{11}{20} \times 0.634$$

$$= 0.594$$

$$\boxed{IG = 0.406}$$

For $x_1 = 11$

x_1	$y = 1$	$y = 2$	entropy
$x_1 < 11$	7	0	0
$x_1 \geq 11$	3	10	0.779

$$\text{Total } E = \frac{13}{20} \times 0.779 = 0.50635$$

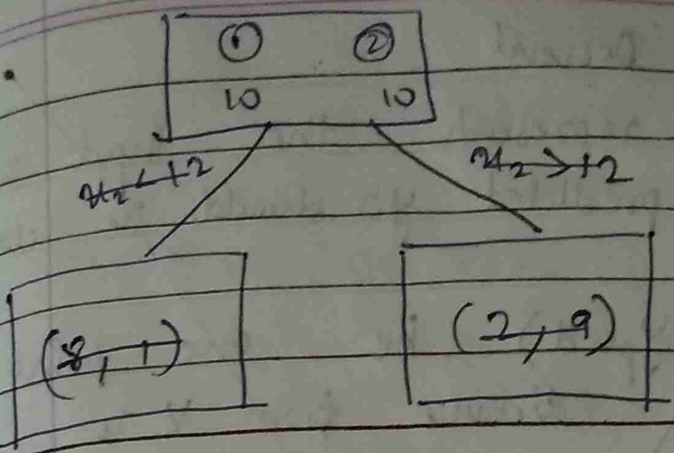
$$IG = 1 - 0.50635 = \boxed{0.49365}$$

For $x_1 = 12$

x_1	$y = 1$	$y = 2$	entropy
$x_1 \leq 12$	8	1	0.503
$x_1 > 12$	2	9	0.634

$$\text{Total } E = \cancel{0.8919}\ 0.594$$

$$IG = 0.406$$



151
352
0.503

0.237
0.447
0.684

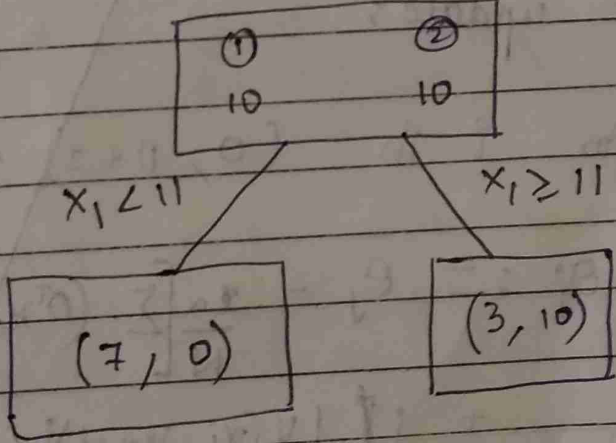
For $x_1 = 13$

	$y=1$	$y=2$	entropy
$x_1 \le 13$	9	2	0.684
$x_1 > 13$	1	8	0.503

$$\text{Total } E = \frac{11}{20} \times 0.684 + \frac{9}{20} \times 0.503$$

$$= 0.6025$$

$$\therefore \boxed{IG = 0.3975}$$



0.488
0.291

4(b) $P(Y = /x_1, x_2) = \frac{P(x_1\, x_2 / y)\, P(y)}{P(x_1, x_2)}$

$= k\, P(x_1/y)\; P(x_2/y)\; P(y)$

Now class conditionals are

1) $P(x_1/y=1) = \sigma(\mu, \sigma) \cdot \frac{k_1}{2}$

$\mu = 8.9$

$s = 6.76$

$P(x_1/y=1) = \frac{k_1}{2}\, \sigma(8.9, 6.76)$

2) $P(x_2/y=1) \Rightarrow$

$(\mu = 9.2 \quad s = 7.228)$

$\Rightarrow \frac{k_2}{2}\, \sigma(9.2, 7.228)$

3) $P(x_1/y=2) \Rightarrow \quad \mu = 14.4$

$s = 4.933$

$\Rightarrow \sigma(14.4, 4.933) \cdot \frac{P}{2}$

4) $P(x_2/y=2) = \sigma(9.9, 16.544)$ $\frac{n}{2}$

$\boxed{k_1 = n = 1}$ $\boxed{\begin{matrix} \mu = 9.9 \\ s = 16.544 \end{matrix}}$

Now for least confidense for x_1 & x_2

$P(y=1/x_1\ x_2) = P(y=2/x_1, x_2)$

for a pair of (x_1, x_2) this should satisfy.

$\sigma(9.2, 7.288)\ \sigma(8.9, 6.76)$

$= \sigma(9.9, 16.544)\ \sigma(14.4, 4.93)$

5 (i) Using Gradient Descent :

$$y = ax + b$$
$$y = \theta^T x = h_\theta(x)$$

* loss function

$$J(\theta) = \sum_{i=1}^{N} (\theta^T x_i - y_i)^2 w_i$$

* weight update

for (i in [0, D+1]):

$$\theta_j := \theta_j - 2\alpha \sum_{i=1}^{N} (\theta^T x_i - y_i) w_i x_i$$

simultaneous update of all the weights

* Repeat the same till max-iter or convergence

Using OLE

$$y = h_\theta(x) =$$
$$\text{error} = (h_\theta(x) - y) = (h_\theta(x) - y)$$

Now

1) W is N × N strictly diagonal matrix with

N(i, i) = weight of ith element colu

2) X is N × D

3) Y is N × 1

(OLS) error = $(y - x\theta)^T \omega (y - x\theta)$

$\Rightarrow (y^T - \theta^T x^T)(\omega^T y - \omega^T x\theta)$

$\Rightarrow y^T \omega^T y - y^T \omega^T x\theta - \theta^T x^T \omega^T y + \theta^T x^T \omega^T x\theta$

$\Rightarrow L = (\omega y)^T y - 2\theta^T x^T \omega^T y + \theta^T x^T \omega^T x\theta$

$\frac{\partial L}{\partial \theta} = 2x^T \omega^T x\theta - 2x^T \omega^T y = 0$

(for optimal sol^n)

$x^T \omega^T x \theta = x^T \omega^T y$

$$\boxed{\theta = (x^T \omega^T x)^{-1} (x^T \omega^T y)}$$

OLE :

5 (ii) $y = h_\theta(x)$ $\quad \theta' = \theta\{1 - 100\}$

$$\text{error} = (y - X\theta)^T (y - X\theta) + (\theta - v)^T (\theta - v)$$

$$= (y^T - \theta^T X^T)(y - X\theta) + (\theta^T - v^T)(\theta - v)$$

$$L = y^T y - y^T X\theta - \theta^T X^T y + \theta^T X^T X \theta$$

$$+ \underbrace{\theta^T \theta}_{\text{scales}} - \theta^T v + v^T v - v^T \theta$$

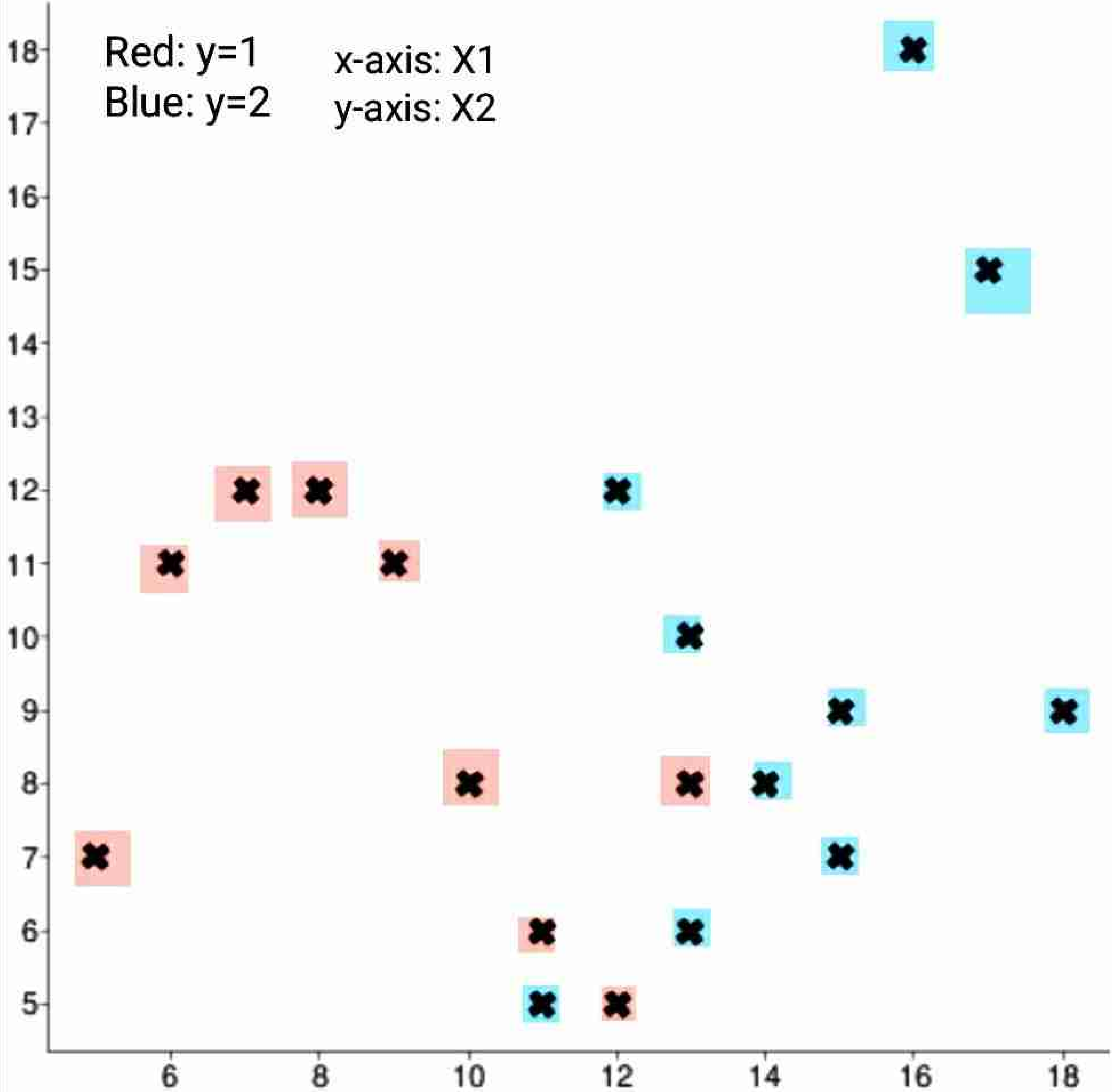
$$\frac{\partial L}{\partial \theta} = 2X^T X\theta - 2X^T y + 2\theta - 2v = 0$$

$$\Rightarrow (X^T X + I)\theta - (X^T y + v) = 0$$

$$\cancel{\theta = (X^T y + v)}$$

$$\boxed{\theta = (X^T X + I)^{-1} (X^T y + v)}$$

Hence the given θ would be closes to the given vector v.


Red: y=1
Blue: y=2
x-axis: X1
y-axis: X2
18
17
16
15
14
13
12
11
10
9
8
7
6
5
6
8
10
12
14
16
18