CS 302

Assignment - 2

180010011

(None)

1 Ans) a) The contingency tables after splitting on attribute A and B are -

	A=T	A=F
+	4	0
−	3	3

	B=T	B=F
+	3	1
−	1	5

The overall entropy before splitting is -

$$E_{orig} = -0.4 \log 0.4 - 0.6 \log 0.6$$

$$= 0.9710$$

The information gain after splitting on A is -

$$E_{A=T} = -\left(\frac{4}{7}\right)\log\left(\frac{4}{7}\right) - \left(\frac{3}{7}\right)\log\left(\frac{3}{7}\right)$$

$$= \boxed{0.9852}$$

$$E_{A=F} = -\left(\frac{3}{3}\right)\log\left(\frac{3}{3}\right) - \left(\frac{0}{3}\right)\log\left(\frac{0}{3}\right)$$

$$= \boxed{0}$$

$$\Delta = E_{orig} - \left(\frac{7}{10}\right)E_{A=T} - \left(\frac{3}{10}\right)E_{A=F}$$

$$= 0.9710 - \left(\frac{7}{10}\right)(0.9852) - \left(\frac{3}{10}\right)(0)$$

$$\boxed{\Delta = 0.2813}$$

∴ Attribute A will be chosen to split the node.

b) The overall Gini index before splitting is

$$G_{orig} = 1 - 0.4^2 - 0.6^2$$

$$= \boxed{0.48}$$

The gain in the Gini index after splitting on A is.

$$G_{A=T} = 1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2$$

$$= \boxed{0.4898}$$

$$G_{A=F} = 1 - \left(\frac{3}{3}\right)^2 - \left(\frac{0}{3}\right)^2$$

$$= \boxed{0}$$

Hence the corresponding gain is equal to

$$G_{orig} - \left(\frac{7}{10}\right) G_{A=T} - \left(\frac{3}{10}\right) G_{A=F}$$

$$= \boxed{0.1371}$$

Similarly, we can compute the gain after splitting on B, which is -

$$G_{orig} - \left(\frac{4}{10}\right) G_{B=T} - \left(\frac{6}{10}\right) G_{B=F}$$

$$= \boxed{0.1633}$$

$\therefore$ Attribute B will be chosen to split the node.

(None)

2 Ans) Let us set the labels for as -1 as - and +1 as +.

We apply a transformation $y = y - 1$ on all $y = 0$.

The points are now $(1,1,-1)$, $(1,-1,1)$, $(-1,1,1)$ and $(-1,-1,-1)$

The given feature spaces is -

$$\phi = (1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2)$$

$$(1,1,-1) \longrightarrow (1, \sqrt{2}, \sqrt{2}, \sqrt{2}, 1, 1)$$

$$(1,-1,1) \longrightarrow (1, \sqrt{2}, -\sqrt{2}, -\sqrt{2}, 1, 1)$$

$$(-1,1,1) \longrightarrow (1, -\sqrt{2}, \sqrt{2}, -\sqrt{2}, 1, 1)$$

$$(-1,-1,-1) \longrightarrow (1, -\sqrt{2}, -\sqrt{2}, \sqrt{2}, 1, 1)$$

we have to minimize $\frac{\|w\|^2}{2}$

with constraint $\left(y_i(w_{x_i} + b) - 1\right)$

$$\frac{\partial L_P}{\partial w} = w - \sum_{1}^{4} \lambda_i y_i x_i = 0$$

$$w = \sum_{1}^{4} \lambda_i y_i x_i$$

$$\frac{\partial L_P}{\partial w} = \sum_{1}^{4} x_i y_i = 0 \qquad \therefore \sum_{1}^{4} \lambda_i y_i = 0$$

$$L_D = \sum_{1}^{4} \lambda_i + \frac{1}{2} \sum_{1,1}^{n,n} \lambda_i y_i \lambda_j y_j x_i x_j - \sum_{1}^{n} \lambda_i y_i \sum_{1}^{n} (\lambda_j y_j x_j) x_i$$

$$= \sum_{1}^{4} \lambda_i - \frac{1}{2} \sum_{i,j}^{n,n} \lambda_i \lambda_j y_i y_j x_i x_j$$

$$\lambda_1 - \lambda_2 + \lambda_3 - \lambda_4 = 0.$$

using symmetry

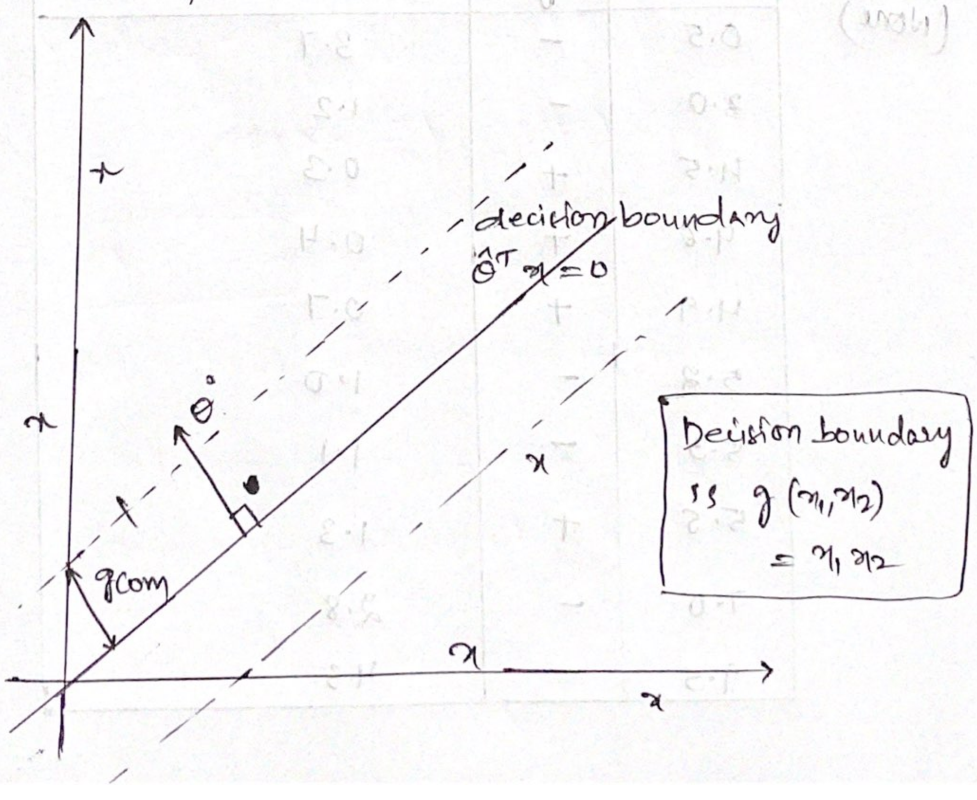
$\lambda_1 = \lambda_2 \qquad \lambda_3 = \lambda_4$

$\lambda_1 = \lambda_3 \qquad \lambda_2 = \lambda_4$

$\therefore \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{8}.$

The maximum margin decision boundary is $g(x_1, x_2) = 0$. (x_1, x_2)

All 4 points are support vectors.



3) (None)

$$f(w) = \frac{\|w\|^2}{2} + C\left(\sum_{i=1}^{N} \xi_i\right)^2$$

Using Lagrange multiplier

$$L_p = \frac{\|w\|^2}{2} + L\left(\sum_{i=1}^{N} \xi_i\right)^2 - \sum \lambda_i \left(y_i(w x_i + b) + \xi_i\right) - \sum_{i=1}^{N} \alpha_i \xi_i$$

$$\frac{\partial L_p}{\partial w} = w - \sum \lambda_i y_i x_i = 0$$

$$w = \sum \lambda_i y_i x_i$$

$$\frac{\partial L_p}{\partial \xi_i} = 2C \sum_{i=1}^{N} \xi_i - \sum \lambda_i - \sum_{i=1}^{N} \alpha_i = 0$$

$$\sum_{i=1}^{N} \xi_i = \frac{\sum \lambda_i + \sum \alpha_i}{2C}$$

$$L_D = \frac{\left(\sum \lambda_i x_i y_i\right)^2}{2} + C\left(\frac{\sum \lambda_i + \sum \alpha_i}{2C}\right)^2 -$$

$$\sum \lambda_i y_i \sum(\lambda_i x_i y_i) x_i - \sum_{i=1}^{N} \alpha_i \left(\frac{\sum \lambda_i + \sum \alpha_i}{2C}\right)$$

4 Ans) (None)

x	y	distance from $x = 4.2$
0.5	−	3.7
3.0	−	1.2
4.5	+	0.3
4.6	+	0.4
4.9	+	0.7
5.2	−	1.0
5.3	−	1.1
5.5	+	1.3
7.0	−	2.8
9.5	−	4.3

a) 1NN $= \{(4.5, +)\}$; Label = +

3NN $= \{(4.5, +), (4.0, +), (4.9, +)\}$

Label = +

5NN $= \{(4.5, +), (4.6, +), (4.9, +), (5.2, -),$
$(5.3, -)\}$; Label = +

9NN $= \{(0.5, -), (3, -), (4.5, +), (4.6, +), (4.9, +),$
$(5.2, -), (5.5, +), (7.0, -)\}$

Label = −

b) 1NN = 0.3+ Label = +

3NN $= (0.3.+) + (0.4.+) + (0.7.+)$

$= 1.4 +$ Label = +.

5NN $= (0.3.+) + (0.4.+) + (0.7.+) + (1.0.-) + (1.1-)$

$= (1.1+) + (2.1.-)$

$\therefore$ −1.0 Label = −

9NN $= (1.0.-) + (3.7.-) + (1.2.-) + (1.3.+)$
$+ (2.8.-)$

$= (8.7.-) + (1.3.+)$

$= (7.4.-)$ Label = −