Assignment 11

Machine Learning

(1) Given points.

$((5, 4), +1)$

$((8, 3), +1)$

$((7, 2), -1)$

$((3, 3), -1)$

$v \geq c + au \rightarrow$ positive

$v < c + au \rightarrow$ negative

If the equation satisfies for the given points and gives the same classes then the equation is true

Let assume $v = 6 - \frac{3u}{5}$

$x_1 : (5, 4) \Rightarrow 4 \geq 6 - \frac{3(5)}{5}$

$4 \geq 3 \quad (T)$

$x_2 : (8, 3) \Rightarrow 3 \geq 6 - \frac{24}{5}$

$3 \geq 1.2 \quad (T)$

$x_3 : (3, 3) \Rightarrow 3 < 6 - \frac{9}{5}$

$3 < 4.2 \quad (T)$

$x_4 : (7, 2) \Rightarrow 2 < 6 - \frac{21}{5}$

$2 < 1.8 \quad (F)$

The equation doesn't satisfy the fourth point.

Now, let us take $v = 6 - \frac{u}{2}$

$x_1 : (5,4) \Rightarrow 4 \geq 6 - \frac{5}{2}$ $\quad 4 \geq 3.5$ (T)

$x_2 : (8,3) \Rightarrow 3 \geq 6 - \frac{8}{2}$ $\quad 3 \geq 2$ (T)

$x_3 : (7,2) \Rightarrow 2 < 6 - \frac{7}{2}$ $\quad 2 < 2.5$ (T)

$x_4 : (3,3) \Rightarrow 3 < 6 - \frac{3}{2}$ $\quad 3 < 4.5$ (T)

The equation satisfies all points.

(2) The line through the points (5,4) & (8,3)

$$m = \frac{3-4}{8-5} = \frac{-1}{3}$$

$$y = mx + c$$

$$y = \left(\frac{-1}{3}\right)x + b$$

$$4 = \frac{-5}{3} + b \Rightarrow b = \frac{17}{3}$$

Equation of line $v = \frac{17}{3} - \frac{u}{3}$

When we move this line until it meets x_3 or x_4, it meets x_3 first & the equation becomes

$$v = \frac{13}{3} - \frac{u}{3}$$

∴ Parallel lines between these lines

$$v = 5 - \frac{u}{3}$$

This must be put in $ax + b = 0$ form

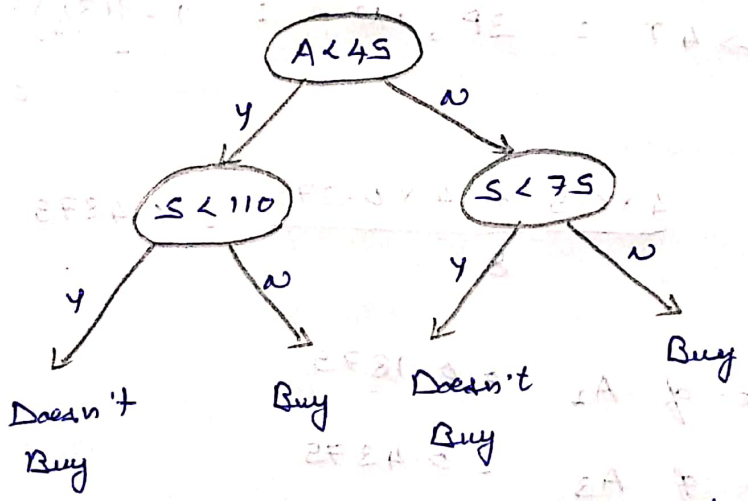
We must scale the vector $w = (1/2, 1)$ & constant $b = -5$, so that if (u, v) is x_1, as x_2, the value would be $+1$ for x_1 & -1 for x_3

Vertical distance b/w boundary & parallel lines through x_1 & x_2 on one hand & through x_3 on other is $2/3$

$\therefore$ Scaling factor $= 3/2$

$\therefore w = \left(\frac{1}{2}, \frac{3}{2}\right)$, $b = \frac{-15}{2}$

(3) Given decision tree,



We need to see any misclassification in the list,

(a) (50, 90)

A is 50, S is 90, This will lead to buy's gold and it is classified correctly

(b) (25, 125)

It is misclassified with tree

(4)

A_1	A_2	A_3	Class Label
F	F	F	+
F	F	T	+
F	T	F	−
F	T	T	−
T	F	F	+
T	F	T	+
T	T	F	+
T	T	T	−

Gini Index of A_1 :

$\rightarrow 4F$: 2P, 2N : $1 - \left[(2/4)^2 + (2/4)^2\right]$ — 0.5

$\rightarrow 4T$: 3P, 1N : $1 - \left[(3/4)^2 + (1/4)^2\right]$ — 0.375

$$= \frac{4 \times 0.5 + 4 \times 0.375}{8} = 0.4375$$

Gini Index of A_2 = 0.1875

Gini Index of A_3 = 0.4375

Total gini index of class labels

$+ : 5 \Rightarrow 1 - \left[\left(\frac{5}{8}\right)^2 + \left(\frac{3}{8}\right)^2\right] = 0.468$

$- : 3$

Goodness of A_1 = Total gini − gini of A_1

= 0.468 − 0.4373 = 0.038

Goodness of A_2 = 0.468 − 0.41875 = 0.0288

Goodness of A_3 = 0.468 − 0.4375 = 0.038

∴ The gradients A_1 and A_3 are equal and A_2 is the best split.

(5) Given, negative points = (1, 4), (3, 3), (3, 1)

positive points = (3, 6), (5, 3)

(a) (4.1, 4.1)

$(1, 4) \Rightarrow \sqrt{(4.1-1)^2 + (4.1-4)^2} = 3.10$

$(3, 3) \Rightarrow \sqrt{(4.1-3)^2 + (4.1-3)^2} = 1.55$

$(3, 1) \Rightarrow \sqrt{(4.1-3)^2 + (4.1-1)^2} = 3.28$

$(3, 6) \Rightarrow \sqrt{(4.1-3)^2 + (4.1-6)^2} = 2.19$

$(5, 3) \Rightarrow \sqrt{(4.1-5)^2 + (4.1-3)^2} = 1.42$ ✓

The minimum distance is for (5, 3) which is positive. So the point (4.1, 4.1) belongs to positive cluster