IS5152 Assignment - I

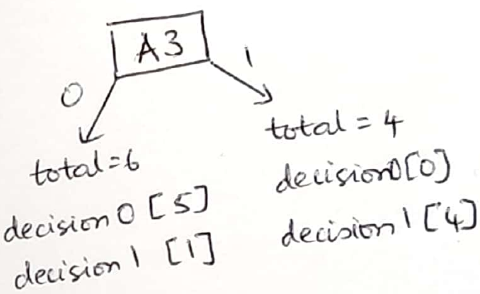
Group 1: Abu Nathew Thoppan (A0178803H)
Balagopal Unnikrishnan (A0178398E)
Gopalakrishan Sai Subramaniam (A0178249N)
Nivedita Valluru Lakshmi (A0178253Y)
Yesupatham Kenneth Rithvik (A0178448M)

1) a) i) No. of responses where decision is yes = 5
No. of responses with decision as no = 5

$Gini_{(decision)} = 1 - (1/2)^2 - (1/2)^2 = 1 - 1/4 - 1/4 = 0.5$

Heterogenity of data using Gini index = 0.5

ii) If A3 is used to split the data,



For subtree 1,

$Gini_{(A3=0)} = 1 - (1/6)^2 - (5/6)^2 = 0.27778$

$Gini\ (A3=1) = 1 - (4/4)^2 = 0$

$Gini\ (A3, decision) = \frac{6}{10} \times 0.2778 + \frac{4}{10} \times 0$

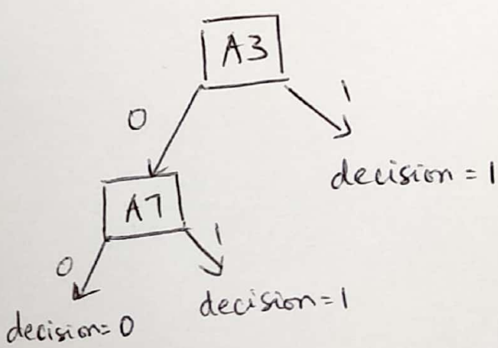
$= 0.16668$

iii) For A3 = 1, the decision is always 1. we only have to find the decision tree split for A3 = 0.

$Gini_{(A7=0, A3=0)} = 1 - (5/5)^2 = 0$

$Gini\ (A7=1, A3=0) = 1 - (1/1)^2 = 0.$

∴ Gini of split using A7 = 0



The final decision tree is shown to the right.

b) From the decision tree, it is obvious that A3 and A7 are enough to make the final decision. It can be noted that A3 ∨ A7 = Decision - "OR" is a linearly seperable function. Hence, the data samples are linearly seperable.

Q-2

(a) mis-classification index $= 1 - \max_n P_n$

no. of cases $= 15$

+ class (LOAN = YES) $= 7$

-ve class (LOAN = NO) $= 8$

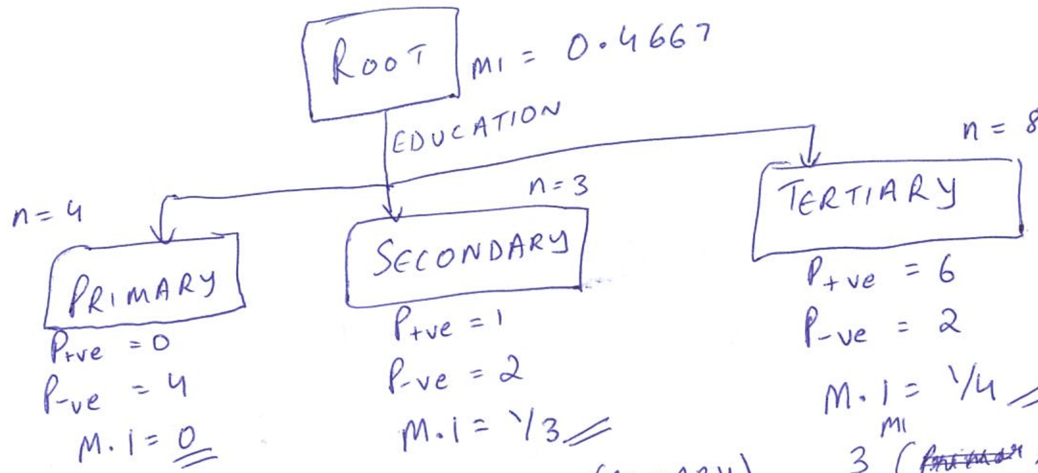
$P_{+ve} = 7/15$ $\qquad P_{-ve} = 8/15$

$\text{Max } P_n = P_{-ve} = 8/15$

$\therefore = 1 - P_n = 7/15$

misclassification index $= 0.4667$ //

(b)



$$\text{GAIN}(S, \text{Education}) = MI(S) - \frac{4}{15} MI(\text{PRIMARY}) - \frac{3}{15} MI(~~\text{Primary}~~ \text{ Secondary}) - \frac{8}{15} MI(\text{tertiary})$$

$$= 0.4667 - \frac{4}{15}(0) - \frac{3}{15}\left(\frac{1}{3}\right) - \frac{8}{15} \times \frac{1}{4}$$

$$= \frac{7}{15} - 0 - \frac{1}{15} - \frac{2}{15} = \frac{4}{15}$$

$= 0.2667$ gain obtained by using education to split.

new $M.I = 0 + \frac{1}{15} + \frac{2}{15} = \frac{3}{15}$

$= 0.2$ //

(c) classification according to rules.

Customer	Predicted	Actual
1	NO	NO
2	NO	NO
3	YES	YES
4	NO	NO
5	YES	YES
6	NO	NO
7	YES	YES
8	YES	YES
9	YES	YES
10	* NO	NO
11	NO	NO
12	* NO	NO
13	NO	YES
14	NO	NO
15	YES	YES

only first matching rule applied.

Confusion Matrix

		Predicted YES	Predicted NO	Total
Actual	YES	6	1	7
	NO	0	8	8
		6	9	

(i) accuracy $= \frac{14}{15} = 93.33\%$

(ii) True positive rate $= \frac{6}{7} = 85.71\%$ | True Positive $= 6$

(iii) True negative $= 8$ | True negative rate $= \frac{8}{8} = 100\%$

3) a) Linear programming:

maximize $60x_1 + 50x_2 + 80x_3 + 160x_4$

subject to the constraints,

$4x_1 + 3x_2 + 6x_3 + 8x_4 \leq 800$

$2x_1 + x_2 + 3x_3 + 4x_4 \leq 300$

$x_1, x_2, x_3, x_4 \geq 0$

b) Dual problem:

minimize: $800u_1 + 300u_2$

subject to the constraints,

$4u_1 + 2u_2 \geq 60$

$3u_1 + u_2 \geq 50$

$6u_1 + 3u_2 \geq 80$

$8u_1 + 4u_2 \geq 160$

$u_1, u_2, \cancel{u_3} \geq 0$

c) We can use the dual problem to solve this LP.

The constraints 1, 3 and 4 are all of the same slope. Hence, we can choose the one that must be satisfied to ensure all 3 are satisfied.

ie, $4u_1 + 2u_2 \geq 60$

$6u_1 + 3u_2 \geq 80$

$8u_1 + 4u_2 \geq 160$ can be further simplified as

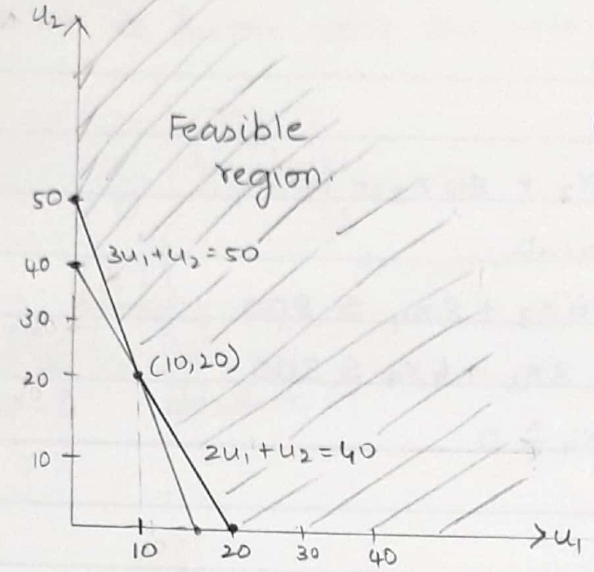
$2u_1 + u_2 \geq 30$

$2u_1 + u_2 \geq 80/3$

$\boxed{2u_1 + u_2 \geq 40}$

If $2u_1 + u_2 \geq 40$ is satisfied, the other two constraints would also be satisfied. Hence, we just use

$2u_1 + u_2 \geq 40$ and $3u_1 + u_2 \geq 50$ to solve this problem.



The optimal solution could be the point of intersection of the two lines and the feasible region is highlighted in the graph.

The point of intersection of the two lines is (10, 20).

Hence, the profit is $800 \times 10 + 300 \times 20$

$= 8000 + 6000 = 14{,}000$

Now, we need to solve for x_1, x_2, x_3 and x_4 given u_1 and u_2.

$$(A^T . u - c)^T . x = 0$$

$$\left(\begin{bmatrix} 4 & 2 \\ 3 & 1 \\ 6 & 3 \\ 8 & 4 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \end{bmatrix} - \begin{bmatrix} 60 \\ 50 \\ 80 \\ 160 \end{bmatrix} \right)^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} 20 \\ 0 \\ 40 \\ 0 \end{bmatrix}^T . \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 0$$

Thus, the equation is $20x_1 + 0 \cdot x_2 + 40 \cdot x_3 + 0 \cdot x_4 = 0$

All $x_i \geq 0$ $\Rightarrow$ x_2 and x_4 are greater than 0 and $x_1 = x_3 = 0$.

$\therefore$ The primal equations are,

$$3x_2 + 8x_4 \leq 800$$

$$x_2 + 4x_4 \leq 300$$

Slack variables s_1 and s_2 are equal to 0 since $s_i \cdot u_i = 0$

and u_1 & u_2 are greater than 0 $\Rightarrow$ $s_1 = s_2 = 0$

$$\therefore \quad 3x_2 + 8x_4 = 800$$

$$x_2 + 4x_4 = 300$$

$\Rightarrow x_2 = 200$

$x_4 = 25$

Thus, to get a maximum profit of 14,000, the company has to produce just 200 chairs and 25 cupboards. Tables and beds should not be produced.

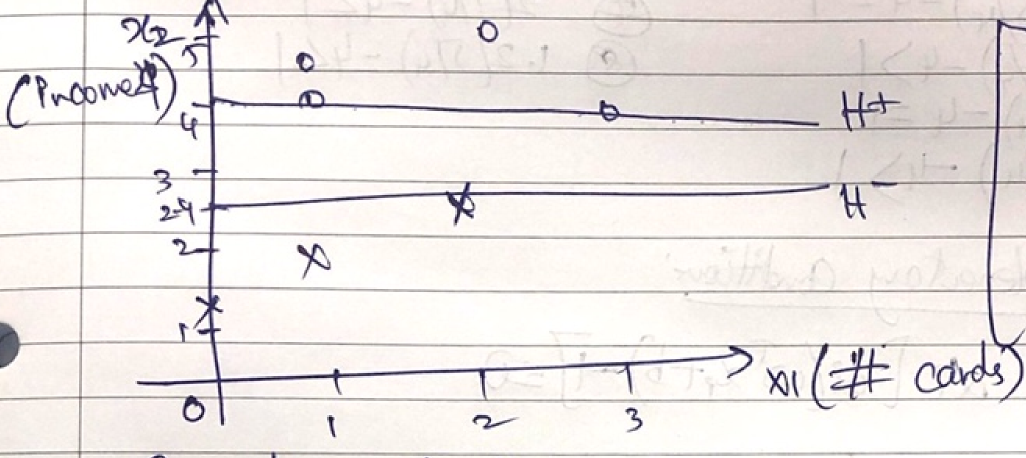
d) One additional hour of carpentry will increase the ~~revenue~~ profit by 10 but one additional hour of varnishing will increase ~~for~~ profit by 20. ($\because u_1 = 10$ & $u_2 = 20$).

Therefore, it is better to choose one additional hour of varnishing to increase the profit.

Date No.

4) a) Plotting the data samples;

Let bad (−1) be represented by X, good (+1) be O.



Maximum margin: $= \dfrac{H^+ - (-H^-)}{2}$

Support vectors:

Point 2 $\Rightarrow 2w_1 + 2.4w_2 + b = -1$

$\Rightarrow 0 + 2.4w_2 + b = -1 \longrightarrow (1)$

Point 4, $\Rightarrow 3w_1 + 4w_2 + b = 1$

$\Rightarrow 0 + 4w_2 + b = 1 \longrightarrow (2)$

Solving (1) & (2), we get $\boxed{w_2 = 5/4,\ b = -4}$.

$\therefore$ The hyperplane eqn: $g(x) = w^T x + b = 0$

$\Rightarrow \boxed{0.x_1 + 5/4\, x_2 - 4 = 0}$

b) To verify whether hyperplane is optimal,

KT necessary conditions: $0w_1 + 2w_2 + b \leq -1$ — (1)

$2w_1 + 2.4w_2 + b \leq -1$ — (2)

$2w_1 + 5w_2 + b \geq 1$ — (3)

$3w_1 + 4w_2 + b \geq 1$ — (4)

$1w_1 + 4.5w_2 + b \geq 1$ — (5)

$1w_1 + 4w_2 + b \geq 1$ — (6)

$1w_1 + 2w_2 + b \leq -1$ — (7)

$0w_1 + 1.2w_2 + b \leq -1$ — (8)

Date No.

Subs. $w_1 = 0, w_2 = 5/4, b = -4$

① : $2(5/4) - 4 < 1$
② : $2.4(5/4) - 4 = -1$
③ : $5(5/4) - 4 > 1$
④ : $4(5/4) - 4 = 1$
⑤ $4.5(5/4) - 4 > 1$
⑥ $4(5/4) - 4 = 1$
⑦ $2(5/4) - 4 < 1$
⑧ $1.2(5/4) - 4 < 1$

KT Complementary Conditions:

$$\alpha_i[d_i(w^T x_i + b) - 1] = 0$$

for non-support vectors, $\alpha_i = 0$
support vectors, $\alpha_i > 0$

$\sum_{i=1}^{8} \alpha_i d_i = 0 \Rightarrow -\alpha_2 + \alpha_4 + \alpha_6 = 0$ (for support vectors)
[Subs. $d_i = -1$ for 2 & $d_i = 1$ for 4 & 6]

$w = \sum_{i=1}^{8} \alpha_i d_i x_i \Rightarrow \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 \\ 5/4 \end{bmatrix} = -\alpha_2 \begin{bmatrix} 2 \\ 2.4 \end{bmatrix} + \alpha_4 \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \alpha_6 \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

$-2\alpha_2 + 3\alpha_4 + \alpha_6 = 0 \longrightarrow$ ⑨
$-2.4\alpha_2 + 4\alpha_4 + 4\alpha_6 = 5/4 \longrightarrow$ ⑩
$-\alpha_2 + \alpha_4 + \alpha_6 = 0 \longrightarrow$ ⑪

⑨ − ⑪ $\Rightarrow -\alpha_2 + 2\alpha_4 = 0 \longrightarrow$ ⑫

~~−4 ⑨ ×4~~ ⑩ − 4 × ⑨ $\Rightarrow 1.6\alpha_2 = 5/4 \Rightarrow \boxed{\alpha_2 = \frac{50}{64}}$

Subs α_2 in ⑫, we get, $\boxed{\alpha_4 = \frac{25}{64}}$

Subs. α_2 & α_4 in ⑨, $\boxed{\alpha_6 = 25/64}$

Because $\alpha_2, \alpha_4, \alpha_6 > 0$, KT complementary conditions are satisfied.

∴ Hyperplane $g(x) = 5/4 x_2 - 4 = 0$ is optimal.