Question 1a

Let $p_1 = Pr(MPG = good)$ and $p_2 = Pr(MPG = bad)$ then we can write the entropy of MPG as follows:

$$H(MPG) = \sum_{i=1}^{2}(-p_i \times log_2 p_i)$$

Since $p_1 = \frac{3}{8}$ and $p_2 = \frac{5}{8}$,

$$H(MPG) = (\frac{-3}{8} \times log_2 \frac{3}{8}) + (\frac{-5}{8} \times log_2 \frac{5}{8}) \simeq 0.95$$

Question 1b

Let X and Y be the random variables representing the Displacement and MPG respectively.

$$H(Y|X) = Pr(X = low)H(Y|X = low) + Pr(X = medium)H(Y|X = medium)Pr(X = high) \times H(Y|X = high)$$

$$= \frac{4}{8}H(\frac{1}{4}, \frac{3}{4}) + \frac{2}{8}H(\frac{1}{2}, \frac{1}{2}) + \frac{2}{8}H(\frac{1}{2}, \frac{1}{2})$$

$$\simeq \frac{4}{8}0.81 + \frac{2}{8} + \frac{2}{8} \simeq 0.9$$

Mutual information between Displacement and MPG can be written as:

$$I(MPG; Displacement) = H(MPG) - H(MPG|Displacement) \simeq 0.95 - 0.9 = 0.05$$

Question 1c

Let X and Y be the random variables representing the Horsepower and MPG respectively.

$$H(Y|X) = Pr(X = low)H(Y|X = low) + Pr(X = medium)H(Y|X = medium)Pr(X = high) \times H(Y|X = high)$$

$$= \frac{2}{8}H(1, 0) + \frac{4}{8}H(\frac{1}{4}, \frac{3}{4}) + \frac{2}{8}H(0, 1)$$

$$\simeq \frac{4}{8}0.81 \simeq 0.4$$

Mutual information between Displacement and MPG can be written as:

$$I(MPG; Horsepower) = H(MPG) - H(MPG|Horsepower) \simeq 0.95 - 0.4 = 0.55$$

Question 1d

Full decision tree is shown in Figure 1. The numbers on the left and right represents the good and bad examples respectively.

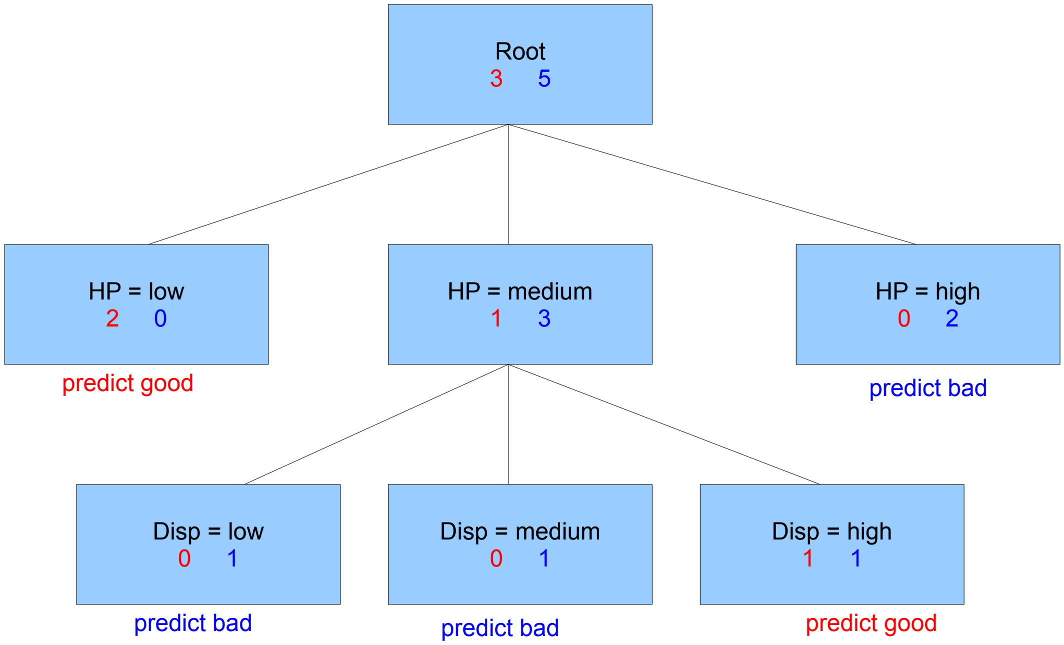

Figure 1: Full decision tree.

Question 2

We can construct a dataset as follows:

A	B	C
0	1	0
1	0	0
0	0	1
1	1	1

I(Y;A) = I(Y;B) = 0. But we can predict the full decision tree perfectly as in Figure 2. The I(Y;A) and I(Y;B) values are verified using the program of Question 4.

Question 3

Since we are in 1D, w will be a scalar value and we can compute the norm of w as follows:

$$||w|| = \sqrt{w'w} = \sqrt{w^2} = w$$

Then we can rewrite the optimization problem as follows:

$$min_{w,b} w$$

$$subject\ to\ y_i(wx_i + b) \geq 1,\ \ i = 1, 2$$

We can plug the x and y values into the problem as follows:

$$min_{w,b} w$$

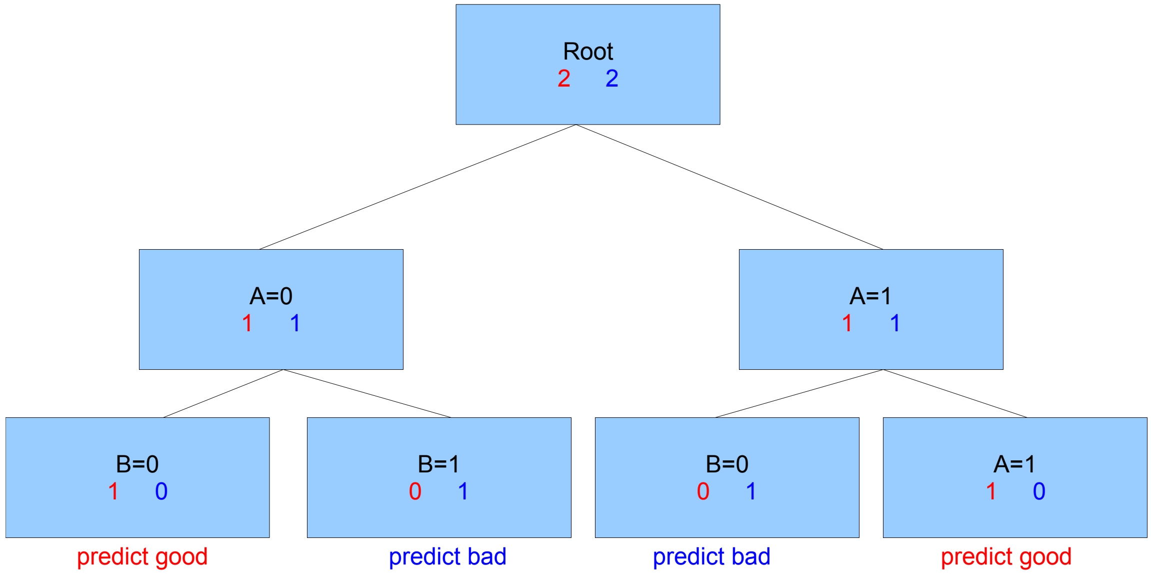

Figure 2: Full decision tree.

$$subject\ to\ -1(w0+b) \geq 1$$

$$subject\ to\ 1(w1+b) \geq 1$$

Which is equal to:

$$min_{w,b} w$$

$$subject\ to\ b \leq -1$$

$$subject\ to\ w+b \geq 1$$

We can rewrite the constraints as follows:

$$subject\ to\ 1-w \leq b \leq -1$$

The lower bound for w is obviously 2 therefore the objective value of the optimization problem is $w = 2$. At this case, $b = -1$.