ASSIGNMENT 2
SECTION A

a) P (Dividend = Yes) = Companies issuing dividends
= 80 % = 0.8

P (Dividends = No) = Companies not issuing dividends
= 100 % − 80 %
= 20 % = 0.2

Mean profit = 10 %
Standard deviation = 36 % } Companies issuing dividends

Mean profit = 0 %
Standard deviation = 36 % } Companies not issuing dividends

Assuming Normal Distribution

$$P(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

→ For companies issuing dividends

$$P(\text{Profit} = 4\% \mid \text{Divident} = \text{Yes}) = \frac{1}{36\sqrt{2\pi}} \exp\left(-\frac{(4-10)^2}{2 \times 36^2}\right)$$

$$= \frac{1}{36\sqrt{2\pi}} \exp\left(\frac{-36}{2592}\right)$$

$$\approx 0.011$$

→ For companies not issuing dividends

$$P(\text{Profit} = 4\% \mid \text{Dividend} = \text{No}) = \frac{1}{36\sqrt{2\pi}} \exp\left(-\frac{(4-0)^2}{2 \times 36^2}\right)$$

$$= \frac{1}{36\sqrt{2\pi}} \exp\left(\frac{-16}{2592}\right)$$

$$\approx 0.0112$$

Using Baye's Theorem,

$$P(\text{Divident} = \text{Yes} \mid \text{Profit} = 4\%) = \frac{P(\text{Profit} = 4\% \mid \text{Divident} = \text{Yes}) \times P(\text{Dividend} = \text{Yes})}{\ldots}$$

$P(\text{Profit} = 4\%)$

→ $P(\text{Profit} = 4\%) = P(\text{Profit} = 4\% \mid \text{Dividend} = \text{Yes}) \times P(\text{Dividend} = \text{Yes})$

$+ P(\text{Profit} = 4\% \mid \text{Dividend} = \text{No}) \times P(\text{Dividend} = \text{No})$

$$P(\text{Dividend} = \text{Yes} \mid \text{Profit} = 4\%) = \frac{0.011 \times 0.8}{(0.011 \times 0.8) + (0.0112 \times 0.2)}$$

$$= \frac{0.0088}{0.0088 + 0.00224}$$

$$= \frac{0.0088}{0.01104} \approx \boxed{0.797}$$

b) For first split,

$$G(S) = 1 - [P(Y)]^2 - [P(N)]^2$$

$$= 1 - \left(\frac{7}{12}\right)^2 - \left(\frac{5}{12}\right)^2 = 1 - \frac{49}{144} - \frac{25}{144}$$

$$= 0.486$$

$= 0.488$

For Class Time,

$P(\text{Morning} / \text{Yes}) = \frac{3}{4}$ $\quad$ $P(\text{Morning} / \text{No}) = \frac{1}{4}$

$$G(\text{Morning}) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2$$

$$= 1 - \frac{9}{16} - \frac{1}{16} = 0.375$$

$P(\text{Noon} / \text{Yes}) = \frac{2}{4} = \frac{1}{2}$ $\quad$ $P(\text{Noon} / \text{No}) = \frac{2}{4} = \frac{1}{2}$

$$G(\text{Noon}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2} = 0.5$$

$P(\text{Afternoon} / \text{Yes}) = \frac{2}{4} = \frac{1}{2}$ $\quad$ $P(\text{Afternoon} / \text{No}) = \frac{2}{4} = \frac{1}{2}$

$$G(\text{Afternoon}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2} = 0.5$$

$$G(\text{Class Time}) = \frac{4}{12} \times 0.375 + \frac{4}{12} \times 0.5 + \frac{4}{12} \times 0.5$$

$$= 0.125 + 0.166 + 0.166$$

$$= 0.458$$

$$IG(\text{Class Time}) = G(S) - G(\text{Class Time})$$

$$= 0.486 - 0.458$$

$$= 0.028$$

For Had Proper Sleep,

$$P(Y/Y) = \frac{6}{6} = 1 \qquad P(Y/N) = 0$$

$$G(Y) = 1 - 1^2 - 0^2 = 0$$

$$P(N/Y) = \frac{1}{6} \qquad P(N/N) = \frac{5}{6}$$

$$G(N) = 1 - (1)^2 - (5)^2$$

$$= 1 - \frac{1}{36} - \frac{25}{36} = 0.277$$

$$G(\text{Had Proper sleeps}) = 0 \times \frac{6}{12} + 0.277 \times \frac{6}{12}$$

$$= 0.1385$$

$$IG(\text{Had proper sleeps}) = G(S) - G(\text{Had Proper Sleeps})$$

$$= 0.486 - 0.1385$$

$$= 0.3745$$

For Weather,

$$P(\text{cool}/Y) = \frac{4}{5} \qquad P(\text{Cool}/N) = \frac{1}{5}$$

$$G(\text{Cool}) = 1 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2$$

$$= 1 - \frac{16}{25} - \frac{1}{25} = 0.32$$

$P(\text{Rainy}, Y) = \frac{0}{2} = 0$ $\qquad$ $P(\text{Rainy} / N) = 1$

$G(\text{Rainy}) = 1 - 0^2 - 1^2 = 0$

$P(\text{Hot} / Y) = \frac{3}{5}$ $\qquad$ $P(\text{Hot} / \text{No}) = \frac{2}{5}$

$$G(\text{Hot}) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2$$

$$= 1 - \frac{9}{25} - \frac{4}{25} = 0.48$$

$$G(\text{Weather}) = \frac{5}{12} \times 0.32 + \frac{2}{12} \times 0 + \frac{5}{12} \times 0.48$$

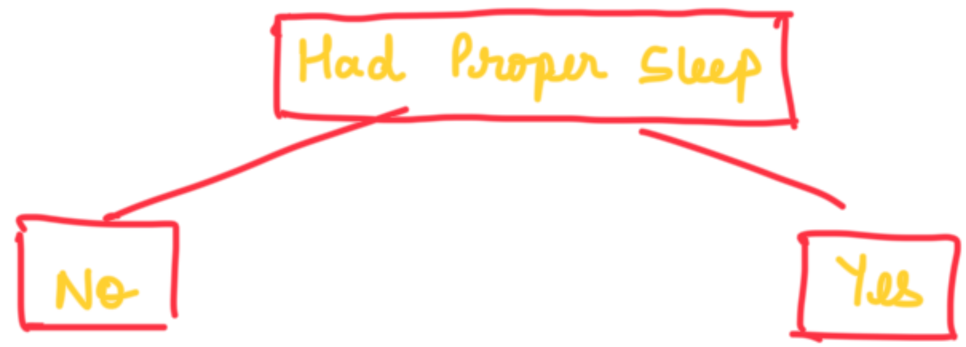
$$= 0.333$$

$$IG(\text{Weather}) = G(S) - G(\text{Weather})$$

$$= 0.486 - 0.333$$

$$= 0.153$$

→ For the first split, Had Proper Sleep is selected.



IG = 0.277   IG = 0 (Leaf Node)

↓

Solve for left half

For Class Time

$$P(\text{Morning} / Y) = \frac{1}{2} \qquad P(\text{Morning} / N) = \frac{1}{2}$$

$$G(\text{Morning}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2}$$

$$P(\text{Noon} / \text{Yes}) = \frac{0}{2} = 0 \qquad P(\text{Noon} / \text{No}) = \frac{2}{2} = 1$$

$$G(\text{Noon}) = 1 - 0^2 - 1^2 = 0$$

$$P(\text{Afternoon}/Y) = \frac{0}{2} = 0 \qquad P(\text{Afternoon}/N) = \frac{2}{2} = 1$$

$$G(\text{Afternoon}) = 1 - 0^2 - 1^2 = 0$$

$$G(\text{Class Time}) = \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times 0$$

$$= \frac{1}{6}$$

$$IG(\text{Class Time}) = G(\text{Left Half}) - G(\text{Class Time})$$
$$= 0.277 - 0.166$$
$$= 0.111$$

For Weather,

$$P(\text{Cool}/Y) = \frac{1}{2} \qquad P(\text{Cool}/N) = \frac{1}{2}$$

$$G(\text{cool}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - \frac{1}{4} - \frac{1}{4} = 0.5$$

$P(Rainy/Y) = \frac{0}{2} = 0$ $\qquad$ $P(Rainy/N) = \frac{2}{2} = 1$

$G(Rainy) = 1 - 0^2 - 1^2 = 0$

$P(Hot/Y) = 0$ $\qquad$ $P(Hot/N) = \frac{2}{2} = 1$

$G(Hot) = 1 - 0^2 - 1^2 = 0$

$$G(Weather) = \frac{1}{3} \times 0.5 + \frac{1}{3} \times 0 + \frac{1}{3} \times 0$$

$$= \frac{1}{6} = 0.166$$

$$IG(Weather) = 0.277 - 0.166$$

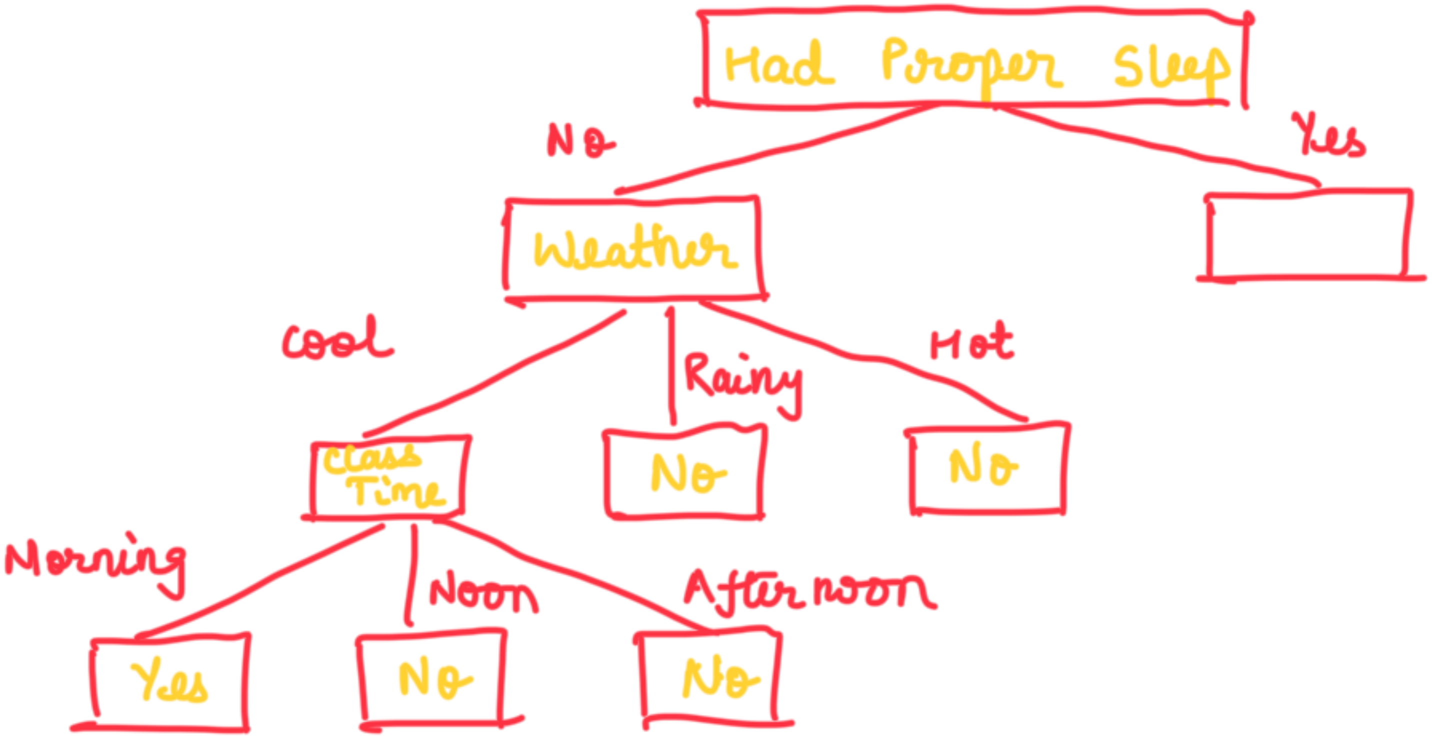
$$= 0.111$$

As you can see, the IG(Class Time) = IG(Weather) = 0.111. We can split with any one of them. So, let's split accross Weather

$G(Left\ (Cool)) = 0.5$

$G(Right\ (Hot)) = G(Middle\ (Rainy)) = 0$ (Leaf Node)

Now, we can split the Left (Cool) Node into Class Time.



c) Margin $= \gamma$

→ Weight vector is updated when there is a classification mistake.

→ Margin mistake occurs if the current hypothesis classifies an example less than $\frac{\gamma}{2}$.

→ Euclidean length is 1.

To Show : This algorithm will halt after a no. of updates that is polynomial in $\frac{1}{\gamma}$. No. of updates is atmost $\frac{8}{\gamma^2}$.

Let w^* be optimal weight vector.
w_t is weight vector at time t.

To calculate projection w onto w^* (optimal vector). we will calculate inner product $= w_t$ and w^*.

After each update $w_t \cdot w^*$ increase by at least $\frac{\gamma}{2}$.

→ Classification mistake : $(w_t + y_i x_i) \cdot w^* \geq w_t \cdot w^* + \gamma$

→ Margin mistake : $(w_t + y_i x_i) \cdot w^* \geq w_t \cdot w^* + \frac{\gamma}{2}$

→ $\|w_t\|$: length of current weight vector.

→ If there is a classification mistake or margin mistake.

$$w_{t+1} \leftarrow w_t + \ln(x)\,x \qquad \ln(x) \in \{-1, 1\}$$

Let w^* be the unit length vector that defines correct linear seperator for each update we measure the progress is the dot product of w_t and w^*.

According to perceptron theorem,

$$w_{t+1} \cdot w^* = w_t \cdot w^* + \gamma$$

Thus after M updates total increase is atleast

$$w_{M+1} \cdot w^* \geq \gamma$$

We split each example x into its orthogonal part and its parallel part (which adds atmost $\frac{\gamma}{2}$ to the length of w_t)

The original algorithm,

$$\|w_{t+1}\|^2 \le \|w_t\|^2 + 1$$

This implies, $\|w_{t+1}\| \le \|w_t\| + \frac{1}{2\|w_t\|}$

The new algorithm,

$$\|w_{t+1}\| \le \|w_t\| + \frac{1}{2\|w_t\|} + \frac{\gamma}{2}$$

After M updates

$$\|w_{M+1}\| \le \frac{2}{\gamma} + \frac{3M\gamma}{4}$$

Using equations : $w_M \cdot w^* \ge \gamma M$

$$\|w_M\| \le \frac{2}{\gamma} + \frac{3M\gamma}{4}$$

Using, $w_M \cdot w^* \le \|w_M\|$

$$\rightarrow \gamma M \le \frac{2}{\gamma} + \frac{3M\gamma}{4}$$

$$\boxed{\Rightarrow \quad M \le \frac{8}{\gamma^2}}$$

d) Spam = 1 = S

Not Spam = 0 = NS

a) $P(S) = \dfrac{\text{No. of smap emails}}{\text{Total no. of emails}} = \dfrac{2}{4} = 0.5$

$P(NS) = \dfrac{2}{4} = 0.5$

For spam,

$P(buy=1/S) = \dfrac{2}{2} = 1$

$P(buy=0/S) = \dfrac{0}{2} = 0$

$P(cheap=1/S) = \dfrac{1}{2} = 0.5$

$P(cheap=0/S) = \dfrac{1}{2} = 0.5$

For not spam,

$$P(buy=1/NS) = \frac{1}{2} = 0.5$$

$$P(buy=0/NS) = \frac{1}{2} = 0.5$$

$$P(cheap=1/NS) = \frac{1}{2} = 0.5$$

$$P(cheap=0/NS) = \frac{1}{2} = 0.5$$

b) The new email contains the word "cheap" but not the word "buy".

$$P(S/buy=0, cheap=1) = \frac{P(buy=0/S) \times P(cheap=1/S) \times P(S)}{P(buy=0, cheap=1)}$$

$$P(NS/buy=0, cheap=1) = \frac{P(buy=0/NS) \times P(cheap=1/NS) \times P(NS)}{P(buy=0, cheap=1)}$$

We will discard the denominator $(P(buy=0, cheap=1))$ because it is constant.

$$P(S / buy=0, cheap=1) = 0 \times 0.5 \times 0.5 = 0$$

$$P(NS / buy=0, cheap=1) = 0.5 \times 0.5 \times 0.5 = 0.125$$

→ The new email will be classified as not spam (Spam=0).

c) The problem with 0 probability arises when we have probabities of a class to be zero in the data, this prevents from accurately predicting the new class because based on prior knowledge, the posterior probability will always be zero. As it can be seen that according to our data if a sentence does not contain the word "buy" then it can never be a spam. Such a problem can be addressed by smoothing techniques Laplace smoothing and m-estimate smoothing.