Theory

1)

Let, $n_+ = 12$ $n_- = 9$

+ve -ve frequencies

$N = n_+ + n_- = 21$

a) $H(Y) = -\left[\frac{n_+}{N}\log_2\left(\frac{n_+}{N}\right) + \frac{n_-}{N}\log_2\left(\frac{n_-}{N}\right)\right]$

$= -\left(\frac{12}{21}\log_2\frac{12}{21} + \frac{9}{21}\log_2\frac{9}{21}\right)$

$= 0.985$

b) $IG(X_i) = H(Y) - H(Y|X_i)$

For $X = x_1$

when, $x_1 = T$:

$n_+ = 7$ $n_- = 1$

$N = 8$

when, $x_1 = F$

$n_+ = 5$ $n_- = 8$

$N = 13$

$H(Y|x_1 = T) = -\left(\frac{7}{8}\log_2\frac{7}{8} + \frac{1}{8}\log_2\frac{1}{8}\right)$

$= 0.54356$

$$H(Y \mid X_1 = F) = -\left(\frac{5}{13}\log_2\frac{5}{13} + \frac{8}{13}\log_2\frac{8}{13}\right)$$

$$= 0.96124$$

$$H(Y \mid X_1) = \frac{8}{21}(0.54356) + \frac{13}{21}(0.96124)$$

$$= 0.80212$$

$$\Rightarrow IG(X_1) = 0.98523 - 0.80212$$
$$= 0.1831$$

For $X = x_2$

when,
$x_2 = T$
$n_+ = 7 \quad n_- = 3$
$N = 10$

when,
$x_2 = F$
$n_+ = 5 \quad n_- = 6$
$N = 11$

$$H(Y \mid x_2 = T) = -\left(\frac{7}{10}\log_2\frac{7}{10} + \frac{3}{10}\log_2\frac{3}{10}\right)$$

$$= 0.88129$$

$$H(y \mid x_2 = F) = -\left(\frac{5}{11}\log_2\frac{5}{11} + \frac{6}{11}\log_2\frac{6}{11}\right)$$

$$= 0.99403$$

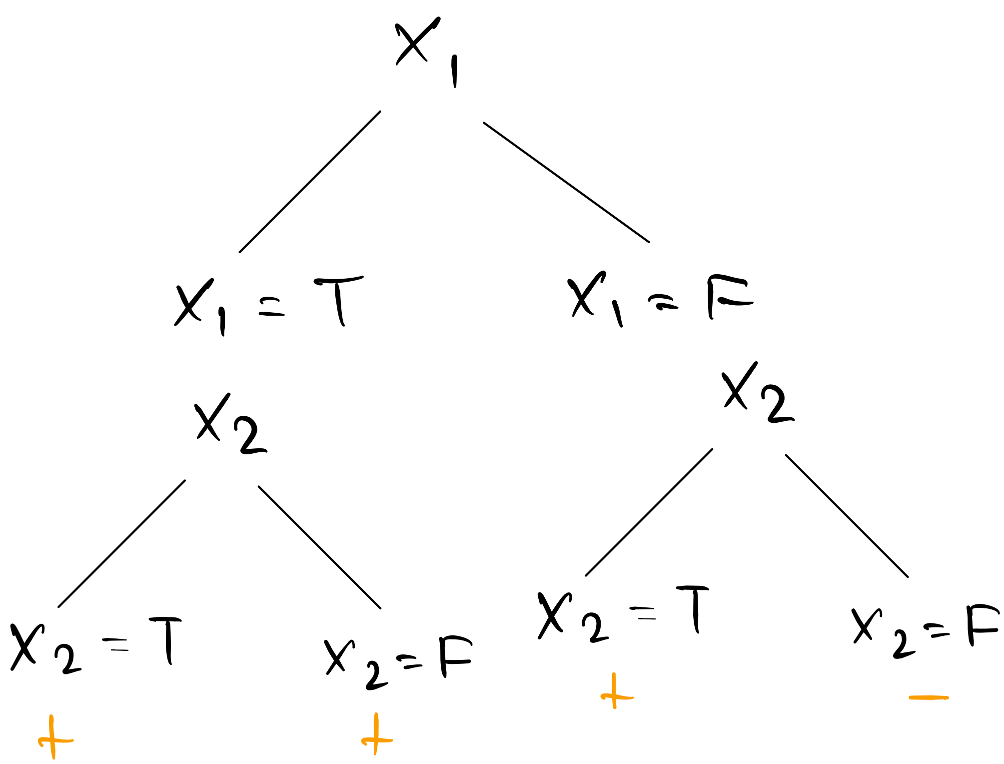
$$H(y \mid x_2) = \frac{10}{21}(0.88129) + \frac{11}{21}(0.99403)$$

$$= 0.94035$$

$$IG(x_2) = 0.98523 - 0.94035$$

$$= 0.04488$$

c)



2)

There are 3 'yes' and 2 'no'

a) $P(Yes) = 0.6 \quad P(No) = 0.4$

b)

	μ	σ
# chars	208	129.88
avg-word-len	4.026	1.1897

$$z = (X - \mu)/\sigma$$

# chars, avglen	z_{chars}, z_{avglen}
216, 5.68	0.0616, 1.395
69, 4.78	-1.0702, 0.6359
302, 2.31	0.7237, -1.4473
60, 3.16	-1.1395, -0.7304
393, 4.2	1.4243, 0.1468

$$\mu_{chars, Yes} = \frac{0.0616 - 1.0702 - 1.1395}{3}$$

$$= -0.716$$

$$\sigma^2_{chars,Yes} = \frac{(0.0616+0.716)^2 + (-1.0702+0.716)^2 + (-1.1395+0.716)^2}{3}$$

$$= 0.3031$$

Similarly,

$\mu_{chars,No} = 1.074 \quad \sigma^2_{chars,No} = 0.1227$

$\mu_{avglen,Yes} = 0.4335 \quad \sigma^2_{avglen,Yes} = 0.7734$

$\mu_{avglen,No} = -0.6503 \quad \sigma^2_{avglen,No} = 0.6352$

c) 242 chars, 4.56 avg length

$$z_{chars} = \frac{242-208}{129.88} = 0.2618$$

$$z_{avglen} = \frac{4.56-4.026}{1.1857} = 0.4504$$

$$f_{chars,Yes}(z) = \frac{\exp\left[-\frac{(z-\mu_{chars,Yes})^2}{\sigma^2}\right]}{\sqrt{2\pi\sigma^2_{chars,Yes}}}$$

Supplying $z = 0.2618$, $\mu = -0.716$,
$\sigma^2 = 0.3031$

in the above function gives:

$$f_{chars, Yes} = 0.2593$$

Similarly for other features and classes

$$f_{chars, No} = 0.0205$$
$$f_{arglen, Yes} = 0.2619$$
$$f_{arglen, No} = 0.7285$$

$$\bar{P}(Yes) = P_{prior}(Yes) \times f_{chars, Yes} \times f_{arglen, Yes}$$
$$= 0.6 \times 0.2593 \times 0.2619$$
$$= 0.04074$$

Similarly,
$$\bar{P}(No) = 0.00598$$

$$P(Yes \mid x) = \bar{P}(Yes) / [\bar{P}(Yes) + \bar{P}(No)]$$
$$= 0.872$$

$P(No \mid X) = 1 - P(Yes \mid X) = 0.128$

$\therefore P(Yes \mid x) > P(No \mid x)$

$\Rightarrow$ Give A!