① Total count = $3+4+4+1+1+3+5=21$

Total +ve = 12   +ve

Total -ve = 9

ⓐ Sample entropy = $-\sum_{i=1}^{N} (-P(v_i)\log_n P(v_i))$

$N=2$

$\Rightarrow \left(-\dfrac{12}{21}\log\dfrac{12}{21}\right) + \left(-\dfrac{9}{21}\log\dfrac{9}{21}\right)$

$= 0.9852$

ⓑ $IG = $ Entropy $-$ average entropy

ⓑ   Entropy 0.9852

average entropy =)

$X_1 \; T = 8$

$X_2 \; T = 10$

$X_1 \; F = 13$

$X_2 \; F = 11$

IG for $X_1 \Rightarrow \dfrac{8}{21} \cdot H\left(\dfrac{7}{8}, \dfrac{1}{8}\right) = \dfrac{8}{21}\left(-\dfrac{7}{8}\log_2\dfrac{7}{8} - \dfrac{1}{8}\log_2\dfrac{1}{8}\right)$

         $= 8.2071$

for $X_1 \Rightarrow \dfrac{13}{21} \cdot H\left(\dfrac{5}{13}, \dfrac{8}{13}\right) = \dfrac{13}{21}\left(-\dfrac{5}{13}\log_2\dfrac{5}{13} - \dfrac{8}{13}\log_2\dfrac{8}{13}\right)$

         $= 0.5951$

Average entropy for $X_1 = 0.2071 + 0.5951$

         ≤ 0.8022

for x_2 T $\Rightarrow \frac{10}{21} \cdot H\left(\frac{7}{10}, \frac{3}{10}\right) = \frac{10}{21} \cdot \left(-\frac{7}{10}\log_2\frac{7}{10} - \frac{3}{10}\log_2\frac{3}{10}\right)$
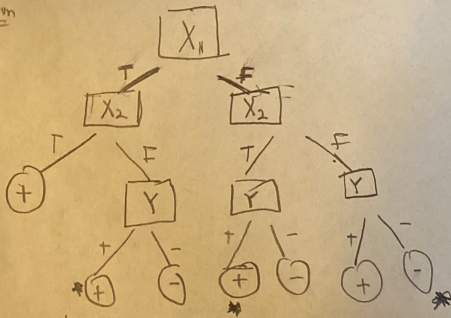
$$= 0.4196$$

for x_2 F $\Rightarrow \frac{11}{21} \cdot H\left(\frac{5}{11}, \frac{6}{11}\right) = \frac{11}{21}\left(-\frac{5}{11}\log_2\frac{5}{11} - \frac{6}{11}\log_2\frac{6}{11}\right)$

$$= 0.5206$$

Average IG of x_2 = $\frac{0.4196 + 0.5206}{}$
entropy

$$= 0.9402$$

Information gain for $X_1 = 0.9852 - 0.8022$

$$= 0.183$$

IG for $X2 = 0.9852 - 0.9402$

$$= 0.045$$

solution

given mode,
the above
is chosen

* C for when mode is used

(2a)

$P(A = Yes) = \frac{3}{5} = 0.6$

$P(A = No) = \frac{2}{5} = 0.4$

(b) standardizing the features,

# of chars = 208
mean

# of chars = 145.22
std

Average word = 4.026
length mean

Std average = 1.3256
word
length

Gaussian parameter

Normalized # of chars	Avg word length	Gave an A
0.0551	1.2477	Yes
-0.9572	0.5688	Yes
0.6473	-1.2445	No
-1.0191	-0.6533	Yes
1.2739	0.1313	No

Yes mean	-0.6464	0.3877
std Yes	0.6033	0.9633
No mean	0.9606	-0.5816
No std	0.4431	1.0082

$= \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_k - \mu_i)^2}{2\sigma_i^2}}$

(c) essay ⇒
Avg word length = 4.562
# of characters = 242
standardize this

Avg word length	No. of characters
0.2169	0.3691

for Yes probability of an A,

$\frac{3}{5} \times 0.2312 \times 0.4141$
↑
from
gaussian
implementation
$= 0.0574$

for not getting an A,

$\frac{2}{5} \times 0.2347 \times 0.2457$
$= 0.0232$.

Since the P(A) is greater, then we say given the
test data above, the essay would get given the
an A.