Approach	Assumptions	Advantages	Disadvantages
Naïve Bayes It is suitable b/c features could be more specified when selected and your model would not be overfitted. would be more general version	The primary assumptions that are made with the Naive Bayes model are that the features are independent of each other. For the email example, we would assume that none of inputs relate to each other	In this particular example. The advantages of using the Naive Bayes model would be that the complexity is greatly reduced in this example. Since there is not as much computation needed.	The disadvantages of using Naive Bayes would be essentially that it is too simple compared to the Bayesian Network model. (This is equivalent to being an underfitted model).
Bayesian Networks Bayesian Networks would be more suitable b/c detecting an email as spam would mean that you would need to connect features together. For instance, the sentence structure and metadata of the email could be interconnected.	The assumption that is made w/ the Bayesian Network model is that the features are depending on each other. W/regards to the example, this means features would be dependent on each other.	The advantage of the Bayesian Network would be that it is a complete model. You could also answer probabalistic queries. What this means is that in your query, we would be able to connect features together.	The disadvantages of the Bayesian network would be that it is computationally expensive and could overfit the model. - Being both overfit and computationally expensive is a disadvantage.

Question 5:

Test (feature)	Has Flu (label C1)	Healthy (label C2)
x_1 → Positive	Ø.8	Ø.1
x_2 → Negative	Ø.2	Ø.9

↑ Test Results ↑

$P(\text{Healthy}) = Ø.25$

$P(\text{Flu}) = 1 - P(\text{Healthy}) = 1 - 0.25$

$P(\text{Flu}) = 0.75$

Build a naive Baye's classifier using the Discriminant Function. Can this classifier predict if a patient is healthy or not-healthy based on the test results? (eg. can a positive test result indicate w/ high confidence that a patient has the flu?)

First, we handle the healthy case:

$$P(\text{healthy} \mid x_1, x_2) \Rightarrow P(\text{healthy}) \cdot \prod_{i=1}^{n} P(x_i \mid \text{healthy}) \cdot P(x_2 \mid \text{healthy})$$

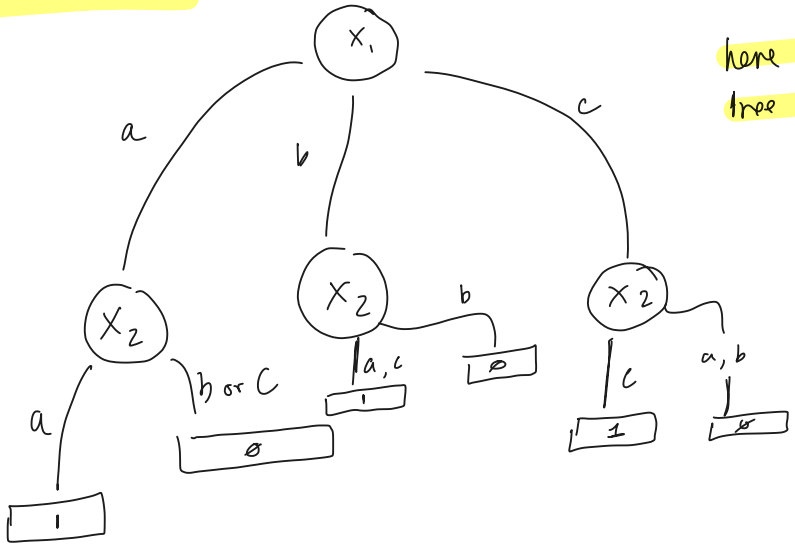
$$Ø.25 \cdot (0.1) \cdot (0.9) = 0.0225$$

Then, we handle the negative case:

$$P(\text{flu} \mid x_1, x_2) \Rightarrow P(\text{flu}) \cdot \prod_{i=1}^{n} P(x_i \mid \text{flu}) \cdot P(x_2 \mid \text{flu})$$

$$Ø.75\ (Ø.8)(Ø.2) = 0.12$$

Q6 → Decision Tree:



here is a decision tree that I constructed

Explanation:

First, we start w/ the fact that X_1 can take any of the values, a, b, or C. After it takes A, B, or C, then, we go into the X_2 node. When we go into X_2, then we go into either 1 or ∅ depending on certain values.

Now, we calculate the entropy of our 'y' values

we have a total of 12 y-values.

6 can be → '0'

6 can go to → '1'

Thus, the $P(y=0) = 0.5$]

and, the probability of $P(y=1) = 0.5$]

* Hint *

$= \log_2(0.5) = -1$

$H(y=0) = -\left(0.5 \times \log_2(0.5) + 0.5 \log_2(0.5)\right)$

$= 1.0$ this value is both for when the output is 0 and when the output is 1.

Now, we calculate the values for entropy for the first feature X_1.

X_1 will work in the following way:

we can get any of the 3 values, a, b, or C.

we calculate the first $P(0|a) = \frac{2}{4}$, $P(1|a) = \frac{2}{4}$

value b: $P(0|b) = \frac{0}{2}$, $P(1|b) = \frac{2}{2}$

as well as for the value of C: $P(0|c) = \frac{2}{4}$, $P(1|c) = \frac{2}{4}$

Now, we use the Entropy formula for each of these, and we see what we will get.

Entropy for values of X_1:

H = this is our entropy function.

$H(Y \mid X_1 = a) = 1$ $\quad H(Y \mid X_1 = c) = 1$

$H(Y \mid X_1 = b) = 0$, $\quad H(Y \mid X_1) = \frac{4}{12} \times 1 + \frac{2}{12} \times 0 + \frac{4}{12} \times 1.0 = \frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3}$

→ Thus, our final entropy value is $\frac{2}{3}$.

Now, we calculate the entropy value for X_2:

a → $p(0 \mid a) = \frac{1}{4}$, $p(1 \mid a) = \frac{3}{4}$

b → $p(0 \mid b) = \frac{2}{2}$, $p(1 \mid b) = \frac{0}{2}$

c → $p(0 \mid c) = \frac{1}{4}$, $p(1 \mid c) = \frac{3}{4}$

Now, we calculate the final entropies.

$H(Y \mid X_2 = a) = 0.81$, $H(Y \mid X_2 = b) = 0$,

$H(Y \mid X_2 = c) = 0.81$, $H(Y \mid X_2) = \frac{4}{12} \times 0.81 + \frac{2}{12} \times 0 + \frac{4}{12} \times 0.81 =$

0.54.

Thus, the $H(Y \mid X_2)$ will be 0.54.

Now, we calculate the IG for both X_1 and X_2.

$IG(X_1) = H(Y) - H(Y \mid X_1) = 1.0 - \frac{2}{3}$

$= \frac{1}{3}$

$IG(X_2) = H(Y) - H(Y \mid X_2) = 1.0 - 0.54$

$= 0.46$

What this means is that the feature X_2 would be selected first.

Question 2:

First, we do something called calculating the posterior probabilities.

$P(spam) = \frac{2}{5}$ (2 spam files out of 5 total files).

$P(ham) = \frac{3}{5}$ (There are 3 Ham files out of 5 total files)

↗ This is all excluding File C6.

Now, we calculate the words that appear in the spam emails (their probabilities)

$P(Price \mid spam) = \frac{2}{2} \rightarrow 1$

$P(Tax \mid spam) = \frac{1}{2} \rightarrow 0.5$

$P(Price \mid Ham) = \frac{2}{3}$

The values for the probability Table →

$P(Price \mid spam) = \frac{1}{6}$

$P(Ticket \mid spam) = \frac{2}{6}$

$P(Free \mid spam) = \frac{1}{6}$

$P(Tax \mid spam) = \frac{1}{6}$

$P(Puppy \mid spam) = \frac{1}{6}$

Ham values:

$P(Price \mid Ham) = \frac{3}{10}$

$P(Ticket \mid Ham) = \frac{2}{10}$

$P(Puppy \mid Ham) = \frac{3}{10}$

Thus, File 6 would be classified as spam.