

**SCHOOL OF ENGINEERING AND COMPUTER SCIENCE**

**Assignment Cover Sheet**

**Full Name: \_\_\_\_\_\_\_\_\_\_\_\_\_Tao Shi\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_**

**Student ID: \_\_300409943\_\_ \_\_Course: \_\_\_\_\_ COM307 \_ \_\_\_**

**Assignment No.: \_\_\_\_\_3\_\_\_ \_ Due Date: 29 May at 11:59pm**

**Part 1: Reasoning Under Uncertainty Basics**

**1. Create the full joint probability table of X and Y, i.e. the table containing the following four joint probabilities P(X = 0, Y = 0), P(X = 0, Y = 1), P(X = 1, Y = 0), P(X = 1, Y = 1). Also explain which probability rules you used.**

**Answer:**

According to the product rule:

P(X = 0, Y = 0) = P(X = 0) \* P(Y = 0| X = 0) = 0.3 \* 0.3 = 0.09

P(X = 0, Y = 1) = P(X = 0) \* P(Y = 1| X = 0) = 0.3 \* 0.7 = 0.21

P(X = 1, Y = 0) = P(X = 1) \* P(Y = 0| X = 1) = 0.7 \* 0.8 = 0.56

P(X = 1, Y = 1) = P(X = 1) \* P(Y = 1| X = 1) = 0.7 \* 0.2 = 0.14

So the full joint probability table is as follows:

|  |  |  |
| --- | --- | --- |
| X | Y | P(X, Y) |
| 0 | 0 | 0.09 |
| 0 | 1 | 0.21 |
| 1 | 0 | 0.56 |
| 1 | 1 | 0.14 |

**2. If given P(X = 1, Y = 0, Z = 0) = 0.336, P(X = 0, Y = 1, Z = 0) = 0.168, P(X = 0, Y = 0, Z = 1) = 0.036, and P(X = 0, Y = 1, Z = 1) = 0.042, create the full joint probability table of the three variables X, Y and Z. Also explain which probability rules you used.**

**Answer:**

Because Z is independent from X given Y, which means P(X, Z|Y) = P(X|Y) \* P(Z|Y).

According to the product rule:

P(X, Z|Y) = P(X, Z, Y) / P(Y)

→P(X, Z, Y) / P(Y) = P(X|Y) \* P(Z|Y)

→P(X, Y, Z) = P(Y) \* P(X|Y) \* P(Z|Y)

→**P(X, Y, Z) = P(X, Y) \* P(Z|Y)**

P(X = 0, Y = 0, Z = 0) = P(X = 0, Y = 0) \* P(Z = 0|Y = 0) = 0.09 \* 0.6 = 0.054

P(X = 1, Y = 0, Z = 1) = P(X = 1, Y = 0) \* P(Z = 1|Y = 0) = 0.56 \* 0.4 = 0.224

P(X = 1, Y = 1, Z = 0) = P(X = 1, Y = 1) \* P(Z = 0|Y = 1) = 0.14 \* 0.8 = 0.112

P(X = 1, Y = 1, Z = 1) = P(X = 1, Y = 1) \* P(Z = 1|Y = 1) = 0.14 \* 0.2 = 0.028

So the full joint probability table is as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| X | Y | Z | P(X, Y, Z) |
| 0 | 0 | 0 | **0.054** |
| 0 | 0 | 1 | 0.036 |
| 0 | 1 | 0 | 0.168 |
| 0 | 1 | 1 | 0.042 |
| 1 | 0 | 0 | 0.336 |
| 1 | 0 | 1 | **0.224** |
| 1 | 1 | 0 | 0.112 |
| 1 | 1 | 1 | 0.028 |

**3. From the above joint probability table of X, Y and Z:**

**(i) calculate the probability of P(Z = 0) and P(X = 0, Z = 0)**

**Answer:**

According to the sum rule:

P(Z = 0) = P(X = 0, Y = 0, Z = 0) + P(X = 0, Y = 1, Z = 0) + P(X = 1, Y = 0, Z = 0) + P(X = 1, Y = 1, Z = 0) = 0.054 + 0.168 + 0.336 + 0.112 = 0.67

P(X = 0, Z = 0) = P(X = 0, Y = 0, Z = 0) + P(X = 0, Y = 1, Z = 0) = 0.054 + 0.168 = 0.222

**(ii) judge whether X and Z are independent to each other and explain why.**

**Answer:**

P(Z = 0) \* P(X = 0) = 0.67 \* 0.3 = 0.201 != P(X = 0, Z = 0)

So X and Z are not independent to each other.

**4. From the above joint probability table of X, Y and Z:**

**(i) calculate the probability of P(X = 1, Y = 0 | Z = 1)**

**Answer:**

P(Z = 1) = 1 - P(Z = 0) = 0.33

P(X = 1, Y = 0 | Z = 1) = P(X = 1, Y = 0, Z = 1) / P(Z = 1) = 0.224 / 0.33 = 0.67879

**(ii) calculate the probability of P(X = 0 | Y = 0, Z = 0).**

**Answer:**

P(X = 0 | Y = 0, Z = 0) = P(X = 0, Y = 0, Z = 0) / P(Y = 0, Z = 0) = 0.054 / (0.054 + 0.336) = 0.13846

**Part 2: Naive Bayes Method**

**1. the probabilities (P(|c) for each feature i.**

**Answer:**

**2. For each input vector F in the unlabelled set, given the input vector F, the probability P(S|D), the probability P(|D), and the predicted class of the input vector. Here D is an email represented by F, S refers to class spam and refers to class non-spam.**

**Answer:**

**3. The derivation of the Naive Bayes algorithm assumes that the attributes are conditionally independent. Why is this like to be an invalid assumption for the spam data? Discuss the possible effect of two attributes not being independent.**

**Answer:**