**Assignment-4 Due Date: 2/17/2022**

1. Consider the following sample data set for Car Theft.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | Color | Type | Manufactured Origin | Stolen |
| 1 | Red | Sports | Domestic | Yes |
| 2 | Red | Sports | Domestic | No |
| 3 | Red | Sports | Domestic | Yes |
| 4 | Yellow | Sports | Domestic | No |
| 5 | Yellow | Sports | Imported | Yes |
| 6 | Yellow | SUV | Imported | No |
| 7 | Yellow | SUV | Imported | Yes |
| 8 | Yellow | SUV | Domestic | No |
| 9 | Red | SUV | Imported | No |
| 10 | Red | Sports | Imported | Yes |
| 11 | Black | Sports | Domestic | No |
| 12 | Black | Sports | Domestic | No |
| 13 | Black | SUV | Imported | Yes |
| 14 | Black | SUV | Domestic | No |
| 15 | Black | SUV | Domestic | Yes |

Using a Naïve Bayes Classifier train the above data set.

a) For the following data, Classify/Predict whether the car would be stolen or not?

1. Red SUV Domestic
   1. Stolen = 9/245:
      1. Red = 3/7, SUV = 3/7, Domestic = 3/7, Stolen = 7/15
   2. Not Stolen = 1/16
      1. Red = 3/8, SUV = 4/8, Domestic = 5/8, Not Stolen = 8/15
2. Black SUV Imported
   1. Stolen = 4/245
      1. Black = 1/7, SUV = 3/7, Imported = 4/7, Stolen = 7/15
   2. Not Stolen = 1/40
      1. Black = 2/8, SUV = 4/8, Imported = 3/8, Not Stolen = 8/15
3. Yellow Imported
   1. Stolen = 4/21
      1. 1/3 \* 4/7 = 4/21
   2. Not Stolen = 1/5
      1. 2/7 \* 7/10 = 1/5
   3. (4/21) \* (7/17) + (1/5) \* (10/17) = 10/51
   4. (1/5) \* (10/17) / (10/51) = 3/5
   5. 1 – 3/5 = 0.4 Probability of being stolen
4. Black Sports
   1. (2/7 \* 3/7 \* 4/7) \* 7/15 + (2/8 \* 3/8 \*4/8) \* 8/15 = 113/1960
   2. (2/7 \* 3/7 \* 4/7) \* 7/15 / (113/1960) = 0.57

2. You will model a small Bayesian network that represents the relationship between yellow fingers, smoking, cancer, radiation, solar flares, and using a microwave.

In this model, smoking can cause yellow fingers and cancer. Solar flares and making microwave popcorn can cause radiation, and radiation can cause cancer as well.

The prior probability of smoking P(S) is 0.3. The prior probability of solar flares P(F) is 0.8. The prior probability of using the microwave is P(M) is 0.9.

The conditional probability table for radiation is

|  |  |  |
| --- | --- | --- |
| F | M | P(R) |
| 0 | 0 | 0.1 |
| 0 | 1 | 0.2 |
| 1 | 0 | 0.2 |
| 1 | 1 | 0.9 |

The conditional probability table for cancer is

|  |  |  |
| --- | --- | --- |
| S | R | P(C) |
| 0 | 0 | 0.1 |
| 0 | 1 | 0.6 |
| 1 | 0 | 0.3 |
| 1 | 1 | 0.9 |

The conditional probability table for yellow fingers is

|  |  |
| --- | --- |
| S | P(Y) |
| 0 | 0.11 |
| 1 | 0.8 |

P(S=1)=0.3

P(S=0)=1-0.3=0.7

Answer the following questions:

1. Draw the Bayesian network for this example?

Diagram

Description automatically generated

1. What is the probability of cancer?
   1. (0.7 \* 0.9 \* 0.8) + (0.7 \* 0.2 \* 0.8) + (0.7 \* 0.1 \* 0.2) + (0.7 \* 0.9 \* 0.2) + (0.3 \* 0.9 \* 0.8) + (0.3 \* 0.2 \* 0.8) + (0.3 \* 0.1 \* 0.2) + (0.3 \* 0.9 \* 0.2) + (0.7 \* 0.9 \* 0.8) + (0.7 \* 0.2 \* 0.8) + (0.7 \* 0.1 \* 0.2) + (0.7 \* 0.9 \* 0.2) + (0.3 \* 0.9 \* 0.8) + (0.3 \* 0.2 \* 0.8) + (0.3 \* 0.1 \* 0.2) + (0.3 \* 0.9 \* 0.2) = 0.53
2. What is the probability of smoking given cancer?
   1. (0.1)\* (1-0.8) \* (1-0.9) + (0.6) \* (1-0.8) \* (0.9) + (0.3) \* (0.2) \* (1-0.9) + (0.9) \* (0.2) \* (0.9) = 0.618
   2. (0.1) \* (0.7) + (0.9) \* (0.3) = 0.34, 0.3 \* 0.9 / 0.34 = 0.78
3. What is the probability of smoking given cancer and radiation?
   1. 0.9 \* 0.9 \* 0.3 / (0.2 \* 0.8 \* 0.3) + (0.1 \* 0.8 \* 0.7) = 0.28
4. What is the probability of cancer if you never use a microwave?
5. Is this a realistic model for predicting the probability of cancer? Explain how can you improve the model?
   1. I believe this model is not that realistic, as this model doesn’t seem like it accounts for a lot of possibilities like genetics and the like, rather its either “you do or you don’t” variables

Question:1 10 points

Question:2 15 points

Total: 25 points