Assignment - 19

Applications of Bayes' Rule

HARSH SIDDHAPURA
1230169813

04/21/2024

PART I: Conditional Probabilities

1. When it is sunny, the number of rows equals 5.
   If you want an outside picnic and the weather is sunny, you determine how many rows to have which is 2.
   Therefore, P(Outdoor Picnic=Y | Sunny=t).
   P = 2/5
   P = **0.4**

2.
   1. **p(Outdoor Picnic = Yes | Windy=True)**
      No. of rows where Windy is True = 5
      No. of rows where Outdoor Picnic is Yes if Windy is True = 0
      p(Outdoor Picnic = Yes | Windy=True) = 0/5 = **0**

   2. **p(Outdoor Picnic = Yes | Rainy=True)**
      No. of rows where Rainy is True = 4
      No. of rows where Outdoor Picnic is Yes if Rainy is True = 0
      p(Outdoor Picnic = Yes | Windy=True) = 0/4 = **0**

   3. **p(Outdoor Picnic = Yes | Hot=True)**
      No. of rows where Hot is True = 5
      No. of rows where Outdoor Picnic is Yes if Hot is True = 2
      p(Outdoor Picnic = Yes | Windy=True) = 2/5 = **0.4**

   4. **p(Outdoor Picnic = Yes | Rainy=False)**
      No. of rows where Rainy is False = 5
      No. of rows where Outdoor Picnic is Yes if Rainy is False = 4
      p(Outdoor Picnic = Yes | Windy=True) = 4/5 = **0.8**

5. **p(Outdoor Picnic = No | Rainy=False, Windy = True). What is your confidence in this probability? Explain**.

No. of rows where Rainy is False and Windy is True = 1
No. of rows where Outdoor Picnic is Yes if Rainy is False and Windy is True = 1
p(Outdoor Picnic = No | Rainy=False, Windy = True) = 1/1 = **1**

- Extracting the single row where Rainy=False and Windy=True gives us the exact subset to calculate the conditional probability. With only 1 row in this subset, which has Outdoor Picnic=No, the conditional probability is unambiguously 1.
- The data is unambiguous and leaves no room for uncertainty in the calculation. When it is not rainy but windy, the probability of having no outdoor picnic is 100% based on the given information.
- The data supports the conditional probability calculation of p(Outdoor Picnic = No | Rainy=False, Windy = True) = 1 with high confidence, as the underlying data is clear and leaves no room for ambiguity in the result.

6. **p(Outdoor Picnic = No | Rainy=False, Windy = False). What is your confidence in this probability? Explain.**

No. of rows where Rainy is False and Windy is False = 4
No. of rows where Outdoor Picnic is No if Rainy is False and Windy is False = 0
p(Outdoor Picnic = No | Rainy=False, Windy = False) = 0/4 = **0**

- Extracting the 4 rows where Rainy=False and Windy=False, we see that none have Outdoor Picnic=No. This means the conditional probability p(Outdoor Picnic = No | Rainy=False, Windy=False) is 0.
- The data is unambiguous and leaves no room for uncertainty in the calculation. When it is not rainy or windy, the probability of having no outdoor picnic is 0% based on the given information.
- Have high confidence in the probability calculation of p(Outdoor Picnic = No | Rainy=False, Windy=False) = 0, as the underlying data is clear and leaves no room for ambiguity in the result. The data supports this conditional probability with certainty.

7. **p(Outdoor Picnic = No | Rainy=False, Hot = True). What is your confidence in this probability? Explain.**

No. of rows where Rainy is False and Hot is True = 3
No. of rows where Outdoor Picnic is No if Rainy is False and Hot is True = 1
p(Outdoor Picnic = No | Rainy=False, Hot = True) = 1/3 = **0.33**

- Extracting the 3 rows where Rainy=False and Hot=True, we see that 1 of them has Outdoor Picnic=No. This means the conditional probability p(Outdoor Picnic = No | Rainy=False, Hot=True) is 1/3 = 0.33.
- While the data is clear and unambiguous, the small sample size (only 3 relevant rows) introduces some uncertainty in the probability estimate. With a larger dataset, the probability could be different.
- Moderate confidence in the probability calculation of p(Outdoor Picnic = No | Rainy=False, Hot=True) = 0.33. The underlying data is clear, but the small sample size means the estimate may not be as reliable as if there were more data points. With a larger dataset, the confidence in this conditional probability would be higher.

8. **p(Outdoor Picnic = Yes | Rainy=False, Hot = True). What is your confidence in this probability? Explain.**

No. of rows where Rainy is False and Hot is True = 3
No. of rows where Outdoor Picnic is Yes if Rainy is False and Hot is True = 2
p(Outdoor Picnic = Yes | Rainy=False, Hot = True) = 2/3 = **0.66**

- Extracting the 3 rows where Rainy=False and Hot=True shows that 2 have Outdoor Picnic=Yes. This means the conditional probability p(Outdoor Picnic = Yes | Rainy=False, Hot=True) is 2/3 = 0.66.
- While the data is clear and unambiguous, the small sample size (only 3 relevant rows) introduces some uncertainty in the probability estimate. With a larger dataset, the probability could be different.
- Moderate confidence in the probability calculation of p(Outdoor Picnic = Yes | Rainy=False, Hot=True) = 0.66. The underlying data is clear, but the small sample size means the estimate may not be as reliable as if there were more data points. With a larger dataset, the confidence in this conditional probability would be higher.

3.

1. **p(picnic=Yes|Sunny=T, Windy=T, Rainy=F, Hot=F)**
    - p(Sunny=True | picnic=Yes) = 2/4
    - p(Windy=True | picnic=Yes) = 0/4
    - p(Rainy=False | picnic=Yes) = 4/4
    - p(Hot=False | picnic=Yes) = 2/4
    - p(picnic=Yes) = 4/9

p(picnic=Yes | Sunny=True,Windy=True,Rainy=False,Hot=False)

= p(Sunny=True | picnic=Yes) × p(Windy=True | picnic=Yes) × p(Rainy=False | picnic=Yes) × p(Hot=False | picnic=Yes) × p(picnic=Yes)

= 2/4 x 0/4 x 4/4 x 2/4 x 4/9

= 0.5 x 0 x 1 x 0.5 x 0.44

= 0

2. **p(picnic=No|Sunny=T, Windy=T, Rainy=F, Hot=F)**
    - p(Sunny=True | picnic=No) = 3/5
    - p(Windy=True | picnic=No) = 5/5
    - p(Rainy=False | picnic=No) = 1/5
    - p(Hot=False | picnic=No) = 2/5
    - p(picnic=No) = 5/9

p(picnic=No | Sunny=True,Windy=True,Rainy=False,Hot=False)

= p(Sunny=True | picnic=No) × p(Windy=True | picnic=No) × p(Rainy=False | picnic=No) × p(Hot=False | picnic=No) × p(picnic=No)

= 3/5 x 5/5 x 1/5 x 2/5 x 5/9

= 0.6 x 1 x 0.2 x 0.4 x 0.55

= 0.0264

The Naive Bayes' Classifier would predict that weather conditions (Sunny=True, Windy=True, Rainy=False, Hot=False) are unfavorable for an outdoor picnic.

The Naive Bayes' Classifier analysis suggests that the weather conditions of Sunny=True, Windy=True, Rainy=False, and Hot=False are not favorable for an outdoor picnic, as the probability of the "No" class (0.0264) is higher than the probability of the "Yes" class (0).

Part II: Evaluating Evidence using Bayes' Rule

1. Both cases can be discussed below:

   - **Case 1: The probability of developing red spots when having measles is 99%**

     In this case, the probability of developing red spots given that the person has measles, p(E|H), is very high at 0.99.
     According to Bayes' Theorem:
     $p(H|E) \propto p(E|H) * p(H)$
     Where:
     - p(H|E) is the probability of having measles given the person has red spots (the diagnosis)
     - p(E|H) is the probability of having red spots given the person has measles
     - p(H) is the prior probability of having measles

     Since p(E|H) is very high at 0.99, this will increase the overall probability of diagnosing p(H|E). In other words, if a person has red spots, there is a very high likelihood that they have measles. The high value of p(E|H) makes the diagnosis of measles much more likely in this case

   - **Case 2: The probability of developing red spots when having measles is 1%**

     In this hypothetical case, the probability of developing red spots given that the person has measles, p(E|H), is very low at 0.01.
     Applying Bayes' Theorem again:
     - $p(H|E) \propto p(E|H) * p(H)$

     Since p(E|H) is very low at 0.01, this will decrease the overall probability of diagnosing p(H|E). In other words, if a person has red spots, there is a very low likelihood of having measles. The low value of p(E|H) makes the diagnosis of measles much less likely in this case

2. Cases can be discussed below:

   - **Case 1: The probability of contracting measles is extremely low in the general population**

     If the prior probability of having measles, p(H), is extremely low in the general population, this will decrease the overall probability of the diagnosis p(H|E). Since p(H) is extremely low, this will reduce the value of p(H|E), making the diagnosis of measles much less likely even if the person has red spots.

   - **Case 2: The probability of contracting measles is steadily increasing over the years**

     If the prior probability of having measles, p(H), is steadily increasing over the years, this will increase the overall probability of the diagnosis p(H|E). As p(H) increases, it will raise the value of p(H|E), making the diagnosis of measles more likely if the person has red spots.

   - **Case 3: The probability of contracting measles is steadily decreasing over the years**

     If the prior probability of having measles, p(H), is steadily decreasing over the years, this will decrease the overall probability of the diagnosis p(H|E). As p(H) decreases, it will lower the value of p(H|E), making the diagnosis of measles less likely even if the person has red spots.

3. If there are many conditions more prevalent than measles that can also lead to the development of red spots, this will increase the overall probability of observing red spots in the general population, p(E). If many other prevalent conditions can cause red spots, the value of p(E) in the denominator will increase. As a result, the overall probability of the diagnosis p(H|E) will decrease.

   In other words, if red spots can be caused by many different, more common conditions besides measles, then observing red spots becomes less indicative of the patient having measles. The diagnosis of measles becomes less likely in this case. The key point is that the more prevalent alternative causes of the observed evidence (red spots) become, the less that evidence supports the specific hypothesis (the patient has measles). The diagnosis becomes less likely.