

ML_ASSISMENT_13

1. Provide an example of the concepts of Prior, Posterior, and Likelihood.

Sol:

case: Medical researcher investigating a new diagnostic test for a certain disease. The disease in question is relatively rare, occurring in about 1 in 1000 people. We have a group of 1000 individuals to test.

- **Prior:** This is the initial belief or probability before taking any new information into account. In this case, our prior belief about an individual having the disease is $1/1000$ (0.001), based on the known prevalence of the disease in the population.
- **Likelihood:** The likelihood is a measure of how well the data fits a particular hypothesis. In this case, it's the probability of observing a positive test result given that an individual actually has the disease or doesn't have the disease. Let's say our diagnostic test is 95% accurate, meaning it has a 95% chance of correctly detecting the disease if the person actually has it, and a 5% chance of giving a false positive for a healthy individual.
- **Posterior:** The posterior is the updated belief after considering new evidence, in this case, the results of the diagnostic test. You use Bayes' theorem to calculate the posterior probability.

Plugging in the values:

Posterior probability = $(0.95 * 0.001) / (0.95 * 0.001 + 0.05 * 0.999) \approx 0.0185$

So, after observing a positive test result, the updated probability that the individual actually has the disease (posterior) is approximately 0.0185 or 1.85%.

2. What role does Bayes' theorem play in the concept learning principle?

Sol:

- Bayes' theorem helps us update our beliefs about the probability of a hypothesis (or concept) being true based on new evidence or observations.
- This is crucial for making informed decisions and updating our understanding as new information becomes available.
- Bayes' theorem provides a mathematical framework for incorporating prior beliefs and evidence to arrive at more accurate and refined conclusions.
- Bayes' theorem is used in the context of concept learning: **Prior Probability, Likelihood, Posterior Probability**

3. Offer an example of how the Naive Bayes classifier is used in real life.

Sol:

Example of Naive Bayes Classifier in Real Life:

Spam Email Detection: One common real-life application of the Naive Bayes classifier is in spam email detection. In this scenario, the classifier is trained on a dataset containing both spam and non-spam (ham) emails. The features could include words, phrases, or other characteristics found in the email's content or metadata.

The classifier calculates the probabilities of an email being spam or ham based on the frequency of words in the email. The "naive" assumption is that the presence of each word is independent of the presence of other words (even though this assumption might not be true in reality).

4. Can the Naive Bayes classifier be used on continuous numeric data? If so, how can you go about doing it?

Sol:

Yes, the Naive Bayes classifier can be used with continuous numeric data, although it's often more suitable for discrete data like text or categorical features.

When dealing with continuous data, we will need to make some adaptations to the standard Naive Bayes algorithm to handle the numerical values appropriately.

There are different variations of the Naive Bayes classifier that can handle continuous numeric data, such as Gaussian Naive Bayes or Kernel Density Estimation Naive Bayes.

Gaussian Naive Bayes: This version assumes that the numeric features follow a Gaussian (normal) distribution. For each class, it calculates the mean and standard deviation of each feature and then uses the Gaussian probability density function to calculate the likelihood.

Here's a high-level overview of Gaussian Naive Bayes with continuous numeric data:**Data Preparation, Calculate Likelihoods, Calculate Class Probabilities**

5. What are Bayesian Belief Networks, and how do they work? What are their applications? Are they capable of resolving a wide range of issues?

Sol:

Bayesian Belief Networks (BBNs), also known as Bayesian Networks or Probabilistic Graphical Models, are a type of probabilistic graphical model that represents a set of random variables and their conditional dependencies using a directed acyclic graph (DAG).

BBNs combine probability theory and graph theory to model uncertain relationships between variables. The nodes in the graph represent random variables, and the edges represent the conditional dependencies between them.

How They Work:

In a BBN, each node represents a random variable, and the edges between nodes indicate probabilistic relationships.

The conditional probabilities are typically specified in conditional probability tables (CPTs), which show the probability distribution of a node given its parents in the graph. By propagating probabilities through the graph, BBNs can be used to compute the probability distribution of any subset of variables, given evidence observed in the network.

Applications:

BBNs have a wide range of applications across various domains, including:

- **Diagnosis and Decision Support:** BBNs can be used to model medical diagnoses and aid in decision-making by considering symptoms, test results, and medical history.
- **Risk Assessment:** BBNs can model risks in complex systems, such as financial markets, engineering systems, and industrial processes.
- **Natural Language Processing:** BBNs can be employed to handle ambiguity and uncertainty in language processing tasks, such as disambiguation and language generation.
- **Image Recognition:** BBNs can model the relationships between different features in an image and aid in object recognition and scene understanding.

Capability to Resolve Issues:

While BBNs are powerful tools for modeling uncertainty and complex relationships, they have their limitations. BBNs are most effective when the underlying relationships can be represented as a DAG and when the conditional dependencies are known or can be estimated accurately. However, BBNs might struggle with modeling situations involving cyclic dependencies, continuous variables, or when the available data is insufficient to estimate accurate probabilities. In such cases, other modeling techniques or more complex probabilistic models might be more suitable.

6. Passengers are checked in an airport screening system to see if there is an intruder. Let I be the random variable that indicates whether someone is an intruder ($I = 1$) or not ($I = 0$), and A be the variable that indicates alarm ($A = 1$). If an intruder is detected with probability $P(A = 1|I = 1) = 0.98$ and a non-intruder is detected with probability $P(A = 1|I = 0) = 0.001$, an alarm will be triggered, implying the error factor. The likelihood of an intruder in the passenger population is $P(I = 1) = 0.00001$. What are the chances that an alarm would be triggered when an individual is actually an intruder?

Sol:

- $P(A = 1|I = 1) = 0.98$ (Probability of an alarm given an intruder)
- $P(A = 1|I = 0) = 0.001$ (Probability of an alarm given a non-intruder)
- $P(I = 1) = 0.00001$ (Probability of an individual being an intruder)

problem: The problem asks for the probability that an alarm would be triggered when an individual is actually an intruder ($P(A = 1|I = 1)$).

Using Bayes' theorem: $P(A = 1|I = 1) = P(I = 1|A = 1) * P(A = 1) / P(I = 1)$

Where:

- $P(I = 1|A = 1)$ is the probability that an individual is an intruder given that an alarm is triggered. = 0.98
- $P(A = 1)$ is the overall probability of an alarm being triggered. = ?
- $P(I = 1)$ is the probability of an individual being an intruder. = 0.00001

Substituting the given values:

$$P(A = 1|I = 1) = (P(I = 1|A = 1) * P(A = 1)) / P(I = 1)$$
$$P(A = 1|I = 1) = (0.98 * P(A = 1)) / 0.00001$$

We don't have the value of $P(A = 1)$, so unless that value is provided or can be calculated from additional information, we can't calculate the exact chances of an alarm triggering for an intruder.

7. An antibiotic resistance test (random variable T) has 1% false positives (i.e., 1% of those who are not immune to an antibiotic display a positive result in the test) and 5% false negatives (i.e., 1% of those who are not resistant to an antibiotic show a positive result in the test) (i.e. 5 percent of those actually resistant to an antibiotic test negative). Assume that 2% of those who were screened were antibiotic-resistant. Calculate the likelihood that a person who tests positive is actually immune (random variable D).

Sol:

Using Bayes' theorem: $P(D = 1|T = 1) = P(T = 1|D = 1) * P(D = 1) / P(T = 1)$

Where:

- $P(D = 1|T = 1)$ is the probability of being immune given a positive test result. 5%
- $P(T = 1|D = 1)$ is the probability of a positive test result given being immune. 1%
- $P(D = 1)$ is the overall proportion of antibiotic-resistant individuals. 2%
- $P(T = 1)$ is the probability of a positive test result.

Substituting the values

$$= ((1 - 0.05) * 0.02) / ((1 - 0.02) * 0.01 + 0.02 * 0.05)$$
$$= (0.95 * 0.02) / 0.0098 + 0.001$$

$\approx 1.76\%$ chances that a person who tests positive is actually immune.

8. In order to prepare for the test, a student knows that there will be one question in the exam that is either form A, B, or C. The chances of getting an A, B, or C on the exam are 30 percent, 20%, and 50 percent, respectively. During the planning, the student solved 9 of 10 type A problems, 2 of 10 type B problems, and 6 of 10 type C problems.

1. What is the likelihood that the student can solve the exam problem?

Sol:

$$\text{Likelihood} = P(A) * P(\text{solving A}) + P(B) * P(\text{solving B}) + P(C) * P(\text{solving C})$$

$$\text{Likelihood} = 0.30 * 0.90 + 0.20 * 0.20 + 0.50 * 0.60$$

$$\text{Likelihood} = 0.27 + 0.04 + 0.30$$

$$\text{Likelihood} = 0.61$$

2. Given the student's solution, what is the likelihood that the problem was of form A?

Sol:

- $P(A|\text{Solved}) = P(\text{Solved}|A) * P(A) / P(\text{Solved})$

Where:

- $P(A|\text{Solved})$ is the probability that the problem was of form A given that the student solved it.
- $P(\text{Solved}|A)$ is the probability that the student solved the problem given that it was of form A (90% in this case).
- $P(A)$ is the probability of getting problem type A (30%).
- $P(\text{Solved})$ is the probability that the student solved any problem, calculated as the sum of the products of solving each problem type and their respective probabilities.
- $$\begin{aligned} P(\text{Solved}) &= P(\text{Solved}|A) * P(A) + P(\text{Solved}|B) * P(B) + P(\text{Solved}|C) * P(C) \\ &= 0.90 * 0.30 + 0.20 * 0.20 + 0.60 * 0.50 \\ &= 0.27 + 0.04 + 0.30 \\ &= 0.61 \end{aligned}$$
- $$\begin{aligned} P(A|\text{Solved}) &= (0.90 * 0.30) / 0.61 \\ P(A|\text{Solved}) &\approx 0.4426 \end{aligned}$$

9. A bank installs a CCTV system to track and photograph incoming customers. Despite the constant influx of customers, we divide the timeline into 5 minute bins. There may be a customer coming into the bank with a 5% chance in each 5-minute time period, or there may be no customer (again, for simplicity, we assume that either there is 1 customer or none, not the case of multiple customers). If there is a client, the CCTV will detect them with a 99 percent probability. If there is no customer, the camera can take a false photograph with a 10% chance of detecting movement from other objects.

1. How many customers come into the bank on a daily basis (10 hours)?

Sol.

- Given that there is a 5% chance of a customer coming into the bank within each 5-minute time period.
- There are 12 time periods in an hour

you can calculate the expected number of customers per hour as:

Expected Customers per Hour = 5%

so, $(0.05) * 12 = 0.6$ customers

- For 10 hours, the total number of expected customers would be:

Total Expected Customers = Expected Customers per Hour * 10

= $0.6 * 10$

= 6 customers

2. On a daily basis, how many fake photographs (photographs taken when there is no customer) and How many missed photographs (photographs taken when there is a customer) are there?

Sol:

Each 5-minute time period:

- Probability of detecting a customer = 5% (0.05)
- Probability of taking a fake photograph
= $(1 - \text{Probability of detecting a customer}) * \text{Probability of false photograph}$
= $(1 - 0.05) * 0.10$
= $0.95 * 0.10$
= 0.095
- Probability of missed photograph
= $\text{Probability of detecting a customer} * (1 - \text{Probability of false photograph})$
= $0.05 * 0.905$
= 0.045
- Expected Fake Photographs = $\text{Probability of taking a fake photograph} * 120$
= $0.095 * 120$
= 11.4
- Expected Missed Photographs = $\text{Probability of missed photograph} * 120$
= $0.045 * 120$
= 5.4

3. Explain the likelihood that there is a customer if there is a photograph?

Sol:

$$P(\text{Customer}|\text{Photograph}) = (P(\text{Photograph}|\text{Customer}) * P(\text{Customer})) / P(\text{Photograph})$$

Where:

- $P(\text{Customer}|\text{Photograph})$ is the probability that a customer is present given that a photograph is taken.
- $P(\text{Photograph}|\text{Customer})$ is the probability of a photograph being taken if a customer is present.
- $P(\text{Customer})$ is the probability of a customer being present.
- $P(\text{Photograph})$ is the probability of a photograph being taken, which can occur in two cases: either a customer is present and a photograph
- $P(\text{Photograph})$
= $P(\text{Photograph}|\text{Customer}) * P(\text{Customer}) + P(\text{Photograph}|\text{No Customer}) * P(\text{No Customer})$
= $P(\text{Photograph}) = (0.99 * 0.05) + (0.10 * 0.95)$
= $0.0495 + 0.095$
= 0.1445
- $P(\text{Customer}|\text{Photograph}) = (0.99 * 0.05) / P(\text{Photograph})$
= $0.0495 / 0.1445$
= 0.34

10. Create the conditional probability table associated with the node Won Toss in the Bayesian Belief network to represent the conditional independence assumptions of the Naive Bayes classifier for the match winning prediction problem.

Sol:

where "Won Toss" was recorded and you know whether the match was won or lost, you can calculate the probabilities:

- $P(\text{Won Toss} = 0 \mid \text{Match Won} = 0)$: The probability of not winning the toss given that the match was lost.
- $P(\text{Won Toss} = 1 \mid \text{Match Won} = 0)$: The probability of winning the toss given that the match was lost.
- $P(\text{Won Toss} = 0 \mid \text{Match Won} = 1)$: The probability of not winning the toss given that the match was won.
- $P(\text{Won Toss} = 1 \mid \text{Match Won} = 1)$: The probability of winning the toss given that the match was won.