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

A1. Let's consider an example scenario of a medical test to detect a disease.

* Prior probability: The probability of having the disease before any test is performed. Let's say the prevalence of the disease in the population is 5%, then the prior probability of having the disease is 0.05.
* Likelihood probability: The probability of obtaining a positive test result given that the person has the disease or not. Let's say the sensitivity of the test is 90%, i.e., the probability of getting a positive test result if the person has the disease is 0.9, and the specificity is 95%, i.e., the probability of getting a negative test result if the person does not have the disease is 0.95. Then, the likelihood probability of getting a positive test result if the person has the disease is 0.9, and the likelihood probability of getting a negative test result if the person does not have the disease is 0.95.
* Posterior probability: The probability of having the disease given a positive test result. We can calculate this using Bayes' theorem. Let's say a person receives a positive test result, then the posterior probability of having the disease can be calculated as follows:

posterior probability = (likelihood probability \* prior probability) / evidence

where evidence = likelihood probability \* prior probability (using the law of total probability)

plugging in the values, we get:

posterior probability = (0.9 \* 0.05) / ((0.9 \* 0.05) + (0.05 \* 0.05))

which simplifies to:

posterior probability = 0.947

This means that there is a 94.7% chance that the person actually has the disease given a positive test result.

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

A2. Bayes' theorem plays a crucial role in the concept learning principle by allowing us to update our prior beliefs about a hypothesis based on new evidence. In concept learning, we have a set of observations and we want to learn a concept that can accurately classify those observations. Bayes' theorem allows us to update our belief in the concept based on the evidence we observe.

For example, suppose we want to learn a concept that distinguishes between apples and oranges based on their color. Our prior belief might be that red and orange colors are more likely to be associated with apples than oranges. However, as we observe more examples, we may find that oranges can also have a similar shade of red. Bayes' theorem allows us to update our prior belief based on the new evidence and arrive at a more accurate posterior belief about the probability of a particular color being associated with apples or oranges.

Bayes' theorem also helps in selecting the most probable hypothesis from a set of competing hypotheses based on the observed evidence. By calculating the posterior probability of each hypothesis given the observed evidence, we can choose the hypothesis with the highest probability as the most likely explanation for the data. This is an important aspect of concept learning as it allows us to make informed decisions and predictions based on the available data.

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

A3. One example of how the Naive Bayes classifier is used in real life is in email spam filtering. The classifier is trained on a dataset of labeled emails, where the labels indicate whether an email is spam or not. The classifier then uses the text features of incoming emails to calculate the probability of the email being spam or not. If the probability of an email being spam is above a certain threshold, the email is flagged as spam and sent to the spam folder. Otherwise, the email is sent to the inbox. This approach allows for efficient and accurate filtering of large volumes of emails, saving time and improving productivity for users.

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

A4. Yes, the Naive Bayes classifier can be used on continuous numeric data. One way to do this is by assuming a probability distribution for the continuous variables, such as the Gaussian distribution. Then, the mean and variance of each variable can be estimated from the training data. During prediction, the probability density function of each continuous variable is calculated using these estimates, and then these probabilities are combined with the probabilities of the discrete variables using Bayes' theorem.

Another approach is to use a discretization technique to convert the continuous variables into discrete variables. For example, one can divide the range of the variable into several bins or categories and assign a discrete value to each bin. Then, the Naive Bayes classifier can be applied on the resulting discrete variables. However, this approach can lead to loss of information and may not work well if the number of bins is chosen poorly.

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?

A5. Bayesian belief networks, also known as Bayesian networks or causal probabilistic networks, are probabilistic graphical models that represent relationships between variables and their probabilities. They are useful in modeling uncertain knowledge and reasoning under uncertainty. Bayesian networks work by using a directed acyclic graph to represent the dependencies between variables and conditional probability tables to quantify the influence of the variables on each other.

The applications of Bayesian belief networks are diverse and range from medical diagnosis to fraud detection and natural language processing. They are used in various domains such as finance, engineering, and science. Bayesian networks can be used for decision-making, predictive modeling, anomaly detection, and risk assessment. They are also used in recommendation systems, bioinformatics, and image processing.

Bayesian belief networks are capable of resolving a wide range of issues, but they have some limitations. They require the input of prior knowledge or expert opinions to construct the network, and their accuracy depends on the quality of the data and the assumptions made in constructing the model. In addition, the computational complexity of the network increases with the number of variables and the size of the dataset. Despite these limitations, Bayesian belief networks are widely used and have been shown to be effective in solving complex problems.

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 I = 0). 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?

A6. In this scenario, we want to find the probability of an alarm being triggered given that an individual is actually an intruder. This is the conditional probability P(I = 1|A = 1), which can be calculated using Bayes' theorem:

P(I = 1|A = 1) = P(A = 1|I = 1) \* P(I = 1) / P(A = 1)

To find the denominator P(A = 1), we can use the law of total probability:

P(A = 1) = P(A = 1|I = 1) \* P(I = 1) + P(A = 1|I = 0) \* P(I = 0)

Substituting the given probabilities, we get:

P(A = 1) = 0.98 \* 0.00001 + 0.001 \* (1 - 0.00001) = 0.0010098

Now we can calculate P(I = 1|A = 1):

P(I = 1|A = 1) = 0.98 \* 0.00001 / 0.0010098 ≈ 0.0097

So the probability of an alarm being triggered when an individual is actually an intruder is approximately 0.0097, or about 0.97%.

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).

A7. Let's define the events:

* A: a person tests positive
* B: a person is resistant to the antibiotic

We are given:

* P(A|B') = 0.01 (false positive rate)
* P(A'|B) = 0.05 (false negative rate)
* P(B) = 0.02 (prevalence of resistance)

We want to calculate P(B|A), the probability that a person is resistant given that they tested positive for the antibiotic.

We can use Bayes' theorem to calculate this probability:

P(B|A) = P(A|B) \* P(B) / P(A)

We need to calculate P(A), the probability of a positive test result. We can use the law of total probability to calculate this:

P(A) = P(A|B) \* P(B) + P(A|B') \* P(B')

P(B') = 1 - P(B) = 1 - 0.02 = 0.98 (prevalence of non-resistance)

P(A|B) = 1 - P(A'|B) = 1 - 0.05 = 0.95 (true positive rate)

P(A|B') = 0.01 (false positive rate)

P(A) = P(A|B) \* P(B) + P(A|B') \* P(B') = 0.95 \* 0.02 + 0.01 \* 0.98 = 0.0291

Now we can substitute these values into Bayes' theorem:

P(B|A) = P(A|B) \* P(B) / P(A) = 0.95 \* 0.02 / 0.0291 = 0.651

Therefore, the likelihood that a person who tests positive is actually immune is 0.651 or 65.1%.

Top of Form

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?

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

A8. Let event A be the probability of the problem being of type A, event B be the probability of the problem being of type B, and event C be the probability of the problem being of type C. Then,

P(A) = 0.3, P(B) = 0.2, P(C) = 0.5

Let event S be the probability of the student solving the problem, and event SA be the probability of the problem being of type A, given that the student solved it.

1. To find the likelihood that the student can solve the exam problem, we need to calculate P(S), which can be done using the law of total probability:

P(S) = P(S|A)P(A) + P(S|B)P(B) + P(S|C)P(C)

The probability of the student solving the problem given that it is of type A is 9/10, given that it is of type B is 2/10, and given that it is of type C is 6/10. So we can substitute these values to get:

P(S) = (9/10)(0.3) + (2/10)(0.2) + (6/10)(0.5) = 0.63

Therefore, the likelihood that the student can solve the exam problem is 0.63.

1. To find the likelihood that the problem was of form A given the student's solution, we need to use Bayes' theorem:

P(SA) = P(S|A)P(A) = (9/10)(0.3) = 0.27

P(SB) = P(S|B)P(B) = (2/10)(0.2) = 0.04

P(SC) = P(S|C)P(C) = (6/10)(0.5) = 0.3

Now we can apply Bayes' theorem to calculate P(A|S):

P(A|S) = P(SA) / P(S) = 0.27 / 0.63 = 0.429

Therefore, the likelihood that the problem was of form A given the student's solution is 0.429.

Top of Form

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)?

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?

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

A9.

1. In a 5-minute time period, the probability of a customer coming in is 0.05. Therefore, the probability of no customer coming in is 1 - 0.05 = 0.95. In a 10-hour period, there are 120 time periods of 5 minutes each. Using the binomial distribution, we can calculate the expected number of customers coming in over the 10-hour period as:

Expected number of customers = 120 x 0.05 = 6

Therefore, on average, six customers come into the bank on a daily basis.

1. The probability of the CCTV taking a false photograph when there is no customer is 0.1. The probability of the CCTV detecting a customer is 0.99. Therefore, the probability of the CCTV taking a photograph when there is no customer is 0.95 x 0.1 = 0.095, and the probability of the CCTV taking a photograph when there is a customer is 0.05 x 0.01 = 0.0005.

To calculate the number of fake photographs taken in a day, we multiply the expected number of time periods with no customer by the probability of a false photograph:

Expected number of fake photographs = 120 x 0.095 = 11.4

Therefore, on average, 11.4 fake photographs are taken in a day.

To calculate the number of missed photographs (photographs taken when there is a customer), we multiply the expected number of customers by the probability of the CCTV not detecting the customer:

Expected number of missed photographs = 6 x (1 - 0.99) = 0.06

Therefore, on average, 0.06 missed photographs are taken in a day.

1. The probability that there is a customer given that there is a photograph can be calculated using Bayes' theorem:

P(customer | photograph) = P(photograph | customer) x P(customer) / P(photograph)

where P(photograph | customer) is the probability of a photograph being taken given that there is a customer, P(customer) is the prior probability of a customer coming in, and P(photograph) is the probability of a photograph being taken.

P(photograph | customer) = 0.99 P(customer) = 0.05 P(photograph) = P(photograph | customer) x P(customer) + P(photograph | no customer) x P(no customer) P(photograph | no customer) = 0.1 P(no customer) = 0.95

Therefore,

P(photograph) = 0.99 x 0.05 + 0.1 x 0.95 = 0.1045

Plugging these values into Bayes' theorem, we get:

P(customer | photograph) = 0.99 x 0.05 / 0.1045 = 0.474

Therefore, if there is a photograph, the likelihood that there is a customer is 0.474.

Top of Form

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 Nave Bayes classifier for the match winning prediction problem in Section 6.4.4.

A10. Assuming that the Won Toss node has two possible values: Yes and No, and the other features are represented by binary nodes (e.g., Sunny: Yes/No), the conditional probability table for the Won Toss node would be as follows:

| **Won Toss** | **P(Won Toss)** |
| --- | --- |
| Yes | p |
| No | 1 - p |

where p is the probability of winning the toss. The probability distribution of each feature given the value of Won Toss would be computed separately and then combined using the Naive Bayes assumption. For example, the conditional probability table for the Outlook feature given Won Toss is Yes would be:

| **Won Toss** | **Outlook** | **P(Outlook | Won Toss)** |
| --- | --- | --- |
| Yes | Sunny | p1 |
| Yes | Overcast | p2 |
| Yes | Rainy | p3 |

where p1, p2, and p3 are the probabilities of having each value of the Outlook feature given that the team has won the toss. Similarly, the conditional probability table for the Outlook feature given Won Toss is No would be computed as well.