# Assignment 13

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

##### *Answer:*

Sure! Let's say we have a bag of colored marbles. The bag contains 80% blue marbles and 20% red marbles. We are interested in determining the probability of drawing a blue marble from the bag.

In this scenario:

* Prior: The prior probability is our initial belief about the probability of drawing a blue marble from the bag before any new evidence is considered. Here, the prior probability of drawing a blue marble is 0.8 because 80% of the marbles in the bag are blue.
* Likelihood: The likelihood represents the probability of the observed data given a particular hypothesis. In this case, the likelihood would be the probability of drawing a blue marble from the bag, given the prior probability of 0.8.
* Posterior: The posterior probability is the updated probability of drawing a blue marble after considering the new evidence or data. It combines the prior probability and the likelihood to give us an updated belief. To calculate the posterior probability, we would use Bayes' theorem, which states that the posterior probability is proportional to the prior probability multiplied by the likelihood. In this example, the posterior probability of drawing a blue marble would be calculated based on the prior probability of 0.8 and the observed data (evidence).

Let's say we draw a marble from the bag without looking at its color and then observe that it's blue. We can use Bayes' theorem to calculate the updated probability of drawing a blue marble (posterior probability):

Prior probability: P(Blue) = 0.8 Likelihood: P(Blue|Evidence) = 1 (since we observed a blue marble)

Applying Bayes' theorem: P(Blue|Evidence) = (P(Evidence|Blue) \* P(Blue)) / P(Evidence)

P(Evidence|Blue) represents the likelihood, which is 1 in this case because we observed a blue marble. P(Evidence) is the probability of observing a blue marble, which is the sum of the probabilities of observing a blue marble and a red marble:

P(Evidence) = P(Evidence|Blue) \* P(Blue) + P(Evidence|Red) \* P(Red) = 1 \* 0.8 + 0 \* 0.2 = 0.8

Substituting the values back into Bayes' theorem: P(Blue|Evidence) = (1 \* 0.8) / 0.8 = 0.8 / 0.8 = 1

Therefore, the posterior probability of drawing a blue marble after observing a blue marble is 1 or 100%.

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

##### *Answer:*

Bayes' theorem plays a fundamental role in the concept learning principle, particularly in the context of probabilistic inference and updating beliefs based on evidence. The concept learning principle aims to understand how individuals or machines learn and categorize objects or events based on observed data or evidence.

Bayes' theorem provides a mathematical framework for updating beliefs or probabilities based on new evidence. It allows us to calculate the posterior probability of a hypothesis or concept given the prior probability and the likelihood of the observed data.

In the context of concept learning, Bayes' theorem helps in the following ways:

1. Prior Knowledge: Bayes' theorem allows us to incorporate prior knowledge or initial beliefs about concepts or hypotheses before observing any evidence. This prior knowledge can be represented as the prior probability distribution.
2. Evidence Evaluation: When new evidence is observed, Bayes' theorem enables us to assess the likelihood of the evidence given a particular hypothesis or concept. This likelihood function quantifies the probability of observing the evidence under different hypotheses.
3. Belief Update: Bayes' theorem combines the prior probability distribution with the likelihood to compute the posterior probability distribution. This posterior distribution represents the updated beliefs or probabilities after considering the evidence. It provides a framework for updating and revising our beliefs about concepts based on new information.
4. Decision Making: The posterior probability distribution obtained from Bayes' theorem can be used for decision making in concept learning. For example, in classification tasks, the posterior probabilities of different classes can be compared to make informed decisions about the category to which a new object or event belongs.

Overall, Bayes' theorem serves as a foundational principle in concept learning by providing a systematic and mathematically sound approach to incorporating prior knowledge, evaluating evidence, and updating beliefs based on observed data.

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

##### *Answer:*

One example of how the Naive Bayes classifier is used in real life is in email spam filtering. Email providers often employ spam filters to automatically detect and filter out unwanted or malicious emails, saving users' time and protecting them from potential threats.

The Naive Bayes classifier is well-suited for this task because it is efficient, scalable, and can handle large amounts of data. The classifier uses Bayes' theorem to calculate the probability of an email being spam or non-spam based on the presence or absence of certain features or words in the email.

Here's a simplified example of how the Naive Bayes classifier could be used for email spam filtering:

1. Training Phase: During the training phase, a large dataset of pre-labeled emails (spam and non-spam) is used to estimate the probabilities of different features occurring in each class. Features could include specific words, the presence of links or attachments, or even structural characteristics of the email.
2. Feature Extraction: Once the training is complete, the email spam filter extracts relevant features from incoming emails. For example, it might look for specific words or patterns that have been found to be indicative of spam.
3. Probability Calculation: The Naive Bayes classifier calculates the probability of an email being spam or non-spam based on the observed features. It assumes that each feature is conditionally independent of others, which is a simplifying assumption known as the "naive" assumption.
4. Classification: Based on the probabilities calculated, the classifier assigns a label (spam or non-spam) to the email. If the calculated probability of an email being spam is higher than a predefined threshold, the email is classified as spam and filtered out.
5. Continuous Learning: To adapt to changing spam patterns and improve accuracy, the classifier can continue to learn from user feedback. When users mark emails as spam or non-spam, this feedback can be used to update the classifier's knowledge and refine its predictions.

By leveraging the Naive Bayes classifier, email providers can efficiently and effectively filter out a significant portion of spam emails, reducing inbox clutter and improving the overall email experience for users.

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

***Answer:***

Yes, the Naive Bayes classifier can be used with continuous numeric data. However, it requires certain assumptions and additional steps to handle continuous features appropriately. Here's an approach known as Gaussian Naive Bayes for handling continuous numeric data:

1. Data Preparation: Ensure that your dataset contains continuous numeric features. If any categorical features are present, they need to be preprocessed or transformed into numeric representations (e.g., one-hot encoding).
2. Data Distribution Assumption: The Gaussian Naive Bayes classifier assumes that the continuous features in each class are distributed according to a Gaussian (normal) distribution. Therefore, you need to check the distribution of each continuous feature and confirm if it reasonably approximates a Gaussian distribution. You can use statistical tests or visualizations like histograms or Q-Q plots to assess the distribution.
3. Parameter Estimation: For each class, estimate the parameters of the Gaussian distribution associated with each continuous feature. Typically, you need to estimate the mean (µ) and standard deviation (σ) of the feature within each class. Maximum Likelihood Estimation (MLE) is commonly used to estimate these parameters.
4. Probability Calculation: Given a new instance with continuous features, calculate the class conditional probabilities using the estimated Gaussian distribution parameters. For each continuous feature in the instance, calculate the probability density function (PDF) of that feature's value given the estimated mean and standard deviation for the corresponding class.
5. Naive Assumption: The Naive Bayes classifier assumes that the features are conditionally independent, given the class. To incorporate this assumption, multiply the class conditional probabilities of each feature together.
6. Prior Probabilities: Consider the prior probabilities of each class. These can be estimated based on the frequency or proportion of instances belonging to each class in the training dataset.
7. Posterior Probability and Classification: Calculate the posterior probability of the instance belonging to each class using Bayes' theorem, which involves multiplying the class prior probability and the product of the class conditional probabilities. Finally, classify the instance by selecting the class with the highest posterior probability.

It's worth noting that the Gaussian Naive Bayes assumes a specific distribution for continuous features, which may not always hold true in real-world scenarios. In such cases, other variations of Naive Bayes or different classification algorithms might be more appropriate.

##### 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 ?

***Answer:***

Bayesian Belief Networks (BBNs), also known as Bayesian Networks or Probabilistic Graphical Models, are powerful probabilistic models that represent relationships between variables using a directed acyclic graph (DAG) and conditional probability tables (CPTs). BBNs are based on principles of Bayesian probability theory and allow for reasoning under uncertainty.

Here's how Bayesian Belief Networks work:

1. Graphical Structure: A BBN consists of a graphical structure that represents the dependencies between variables. Nodes in the graph represent variables, and directed edges represent probabilistic dependencies between the variables.
2. Conditional Probability Tables (CPTs): Each node in the graph has associated conditional probability tables (CPTs) that define the probabilistic relationships between that node and its parent nodes. The CPTs capture the conditional probabilities of each node given its parents.
3. Probabilistic Reasoning: BBNs use probabilistic reasoning to make inferences and predictions about the variables in the network. By propagating probabilities through the graph using the Bayes' theorem and the network's structure, BBNs can calculate posterior probabilities or make predictions based on observed evidence.
4. Evidence Updating: BBNs allow for evidence updating, where the network can be updated with observed evidence about some variables. By providing evidence for certain nodes, the probabilities of other nodes in the network can be updated accordingly.

Applications of Bayesian Belief Networks:

1. Decision Support Systems: BBNs are widely used in decision support systems, helping users make informed decisions by modeling uncertainties and dependencies between variables.
2. Risk Assessment: BBNs are valuable in risk assessment and management, allowing for the modeling of complex dependencies and uncertainties in various domains such as finance, engineering, and healthcare.
3. Medical Diagnosis: BBNs have been applied to medical diagnosis, combining patient symptoms and medical test results to estimate the likelihood of various diseases or conditions.
4. Predictive Modeling: BBNs can be used for predictive modeling tasks such as predicting customer behavior, machine failure, or market trends by incorporating multiple variables and their dependencies.
5. Fault Diagnosis: BBNs are useful in diagnosing faults in complex systems by modeling relationships between symptoms and potential causes, allowing for efficient fault localization.
6. Natural Language Processing: BBNs have been employed in natural language processing tasks, such as language understanding, machine translation, and information extraction, by capturing probabilistic relationships between words and their meanings.

BBNs are capable of resolving a wide range of issues due to their ability to model complex dependencies and uncertainty in real-world scenarios. However, their effectiveness depends on the quality of the model structure, accurate estimation of conditional probabilities, and availability of sufficient data for training the network.

##### 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 ?

***Answer:***

To find the chances that an alarm would be triggered when an individual is actually an intruder, we need to calculate the conditional probability P(I = 1|A = 1).

Using Bayes' theorem, we can express this probability as:

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

Given: P(A = 1|I = 1) = 0.98 (probability of detecting an intruder) P(A = 1|I = 0) = 0.001 (probability of false alarm for non-intruder) P(I = 1) = 0.00001 (likelihood of an intruder)

To calculate 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)

Given: P(I = 0) = 1 - P(I = 1) = 1 - 0.00001 = 0.99999

Now, let's calculate P(A = 1):

P(A = 1) = P(A = 1|I = 1) \* P(I = 1) + P(A = 1|I = 0) \* P(I = 0) = 0.98 \* 0.00001 + 0.001 \* 0.99999 ≈ 0.0000198 + 0.00099999 ≈ 0.00101979

Now, substituting the values into the Bayes' theorem:

P(I = 1|A = 1) = (P(A = 1|I = 1) \* P(I = 1)) / P(A = 1) = (0.98 \* 0.00001) / 0.00101979 ≈ 0.009615

Therefore, the chances that an alarm would be triggered when an individual is actually an intruder is approximately 0.009615, or about 0.96%.

##### 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) ?

##### *Answer:*

To calculate the likelihood that a person who tests positive is actually immune (D), we need to use Bayes' theorem. Let's define the following variables:

D = Random variable indicating immunity (D = 1 if immune, D = 0 if not immune) T = Random variable indicating test result (T = 1 if positive, T = 0 if negative)

Given:

* False positive rate: P(T = 1|D = 0) = 0.01 (1% false positives)
* False negative rate: P(T = 0|D = 1) = 0.05 (5% false negatives)
* Prevalence of antibiotic resistance: P(D = 1) = 0.02 (2% of those screened are immune)

We want to calculate:

* Likelihood of being immune given a positive test: P(D = 1|T = 1)

Using Bayes' theorem, we can express this probability as:

P(D = 1|T = 1) = (P(T = 1|D = 1) \* P(D = 1)) / P(T = 1)

To calculate P(T = 1), we can use the law of total probability:

P(T = 1) = P(T = 1|D = 0) \* P(D = 0) + P(T = 1|D = 1) \* P(D = 1)

Given: P(D = 0) = 1 - P(D = 1) = 1 - 0.02 = 0.98

Now, let's calculate P(T = 1):

P(T = 1) = P(T = 1|D = 0) \* P(D = 0) + P(T = 1|D = 1) \* P(D = 1) = 0.01 \* 0.98 + 0.95 \* 0.02 = 0.0098 + 0.019 = 0.0288

Now, substituting the values into the Bayes' theorem:

P(D = 1|T = 1) = (P(T = 1|D = 1) \* P(D = 1)) / P(T = 1) = (0.95 \* 0.02) / 0.0288 ≈ 0.066

Therefore, the likelihood that a person who tests positive is actually immune (D) is approximately 0.066, or about 6.6%.

##### 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?

***Answer:***

To calculate the likelihood that the student can solve the exam problem, we need to consider the student's proficiency in solving each type of problem and the probability of encountering each type of problem.

Let's define the following variables:

* P(A) = Probability of encountering a problem of form A = 0.3
* P(B) = Probability of encountering a problem of form B = 0.2
* P(C) = Probability of encountering a problem of form C = 0.5

The student's proficiency can be represented as:

* P(S|A) = Probability of solving a problem given it is of form A
* P(S|B) = Probability of solving a problem given it is of form B
* P(S|C) = Probability of solving a problem given it is of form C

Given:

* The student solved 9 out of 10 type A problems, so P(S|A) = 9/10 = 0.9
* The student solved 2 out of 10 type B problems, so P(S|B) = 2/10 = 0.2
* The student solved 6 out of 10 type C problems, so P(S|C) = 6/10 = 0.6

1. Likelihood that the student can solve the exam problem (P(S)):

P(S) = P(A) \* P(S|A) + P(B) \* P(S|B) + P(C) \* P(S|C) = 0.3 \* 0.9 + 0.2 \* 0.2 + 0.5 \* 0.6 = 0.27 + 0.04 + 0.3 = 0.61

Therefore, the likelihood that the student can solve the exam problem is 0.61, or 61%.

1. Likelihood that the problem was of form A given the student's solution (P(A|S)):

Using Bayes' theorem: P(A|S) = (P(S|A) \* P(A)) / P(S)

We have already calculated P(S) in the previous step, so let's calculate P(A|S):

P(A|S) = (P(S|A) \* P(A)) / P(S) = (0.9 \* 0.3) / 0.61 ≈ 0.4426

Therefore, the likelihood that the problem was of form A given the student's solution is approximately 0.4426, or about 44.26%.

##### 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?

***Answer:***

1. To calculate the number of customers coming into the bank on a daily basis (10 hours), we need to consider the probability of a customer arriving in each 5-minute time period and multiply it by the number of time periods in 10 hours.

Given:

* Probability of a customer arriving in a 5-minute time period: 0.05 (5%)
* Number of 5-minute time periods in 10 hours: 10 hours \* 60 minutes / 5 minutes = 120 time periods

Number of customers coming into the bank on a daily basis = Probability of customer arriving \* Number of time periods = 0.05 \* 120 = 6 customers

Therefore, on a daily basis, approximately 6 customers come into the bank.

1. On a daily basis (10 hours), we can calculate the number of fake photographs (photographs taken when there is no customer) and missed photographs (photographs taken when there is a customer) by considering the probabilities of detection and false detection.

Given:

* Probability of detecting a customer if there is a customer: 0.99 (99%)
* Probability of detecting movement from other objects if there is no customer (false positive): 0.10 (10%)

Number of fake photographs (photographs taken when there is no customer) = Probability of false detection \* Number of time periods = 0.10 \* 120 = 12 photographs

Number of missed photographs (photographs taken when there is a customer) = (1 - Probability of detection) \* Number of customers = (1 - 0.99) \* 6 = 0.01 \* 6 = 0.06 photographs

Therefore, on a daily basis, there are approximately 12 fake photographs and 0.06 missed photographs.

1. The likelihood that there is a customer if there is a photograph can be calculated using Bayes' theorem.

Let's define the following variables:

* P(Customer) = Probability of a customer arriving in a 5-minute time period = 0.05 (5%)
* P(Photograph|Customer) = Probability of a photograph being taken if there is a customer = 0.99 (99%)
* P(Photograph|No Customer) = Probability of a photograph being taken if there is no customer = 0.10 (10%)

We want to calculate:

* P(Customer|Photograph) = Likelihood that there is a customer given there is a photograph

Using Bayes' theorem: P(Customer|Photograph) = (P(Photograph|Customer) \* P(Customer)) / P(Photograph)

To calculate P(Photograph), we can use the law of total probability: P(Photograph) = P(Photograph|Customer) \* P(Customer) + P(Photograph|No Customer) \* P(No Customer)

Given:

* P(No Customer) = 1 - P(Customer) = 1 - 0.05 = 0.95

Now, let's calculate P(Photograph):

P(Photograph) = P(Photograph|Customer) \* P(Customer) + P(Photograph|No Customer) \* P(No Customer) = 0.99 \* 0.05 + 0.10 \* 0.95 = 0.0495 + 0.095 = 0.1445

Now, substituting the values into the Bayes' theorem:

P(Customer|Photograph) = (P(Photograph|Customer) \* P(Customer)) / P(Photograph) = (0.99 \* 0.05) / 0.1445

**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…**

***Answer:***

To create the conditional probability table (CPT) associated with the node "Won Toss" in the Bayesian Belief network for the Naive Bayes classifier, we need to specify the conditional probabilities of "Won Toss" given the other variables in the network. Assuming we have two features, "Weather" and "Venue," and the target variable "Won Toss," we can construct a CPT as follows:

css

| Weather | Venue | P(Won Toss = Yes) | P(Won Toss = No) |

|-------------|---------|------------------|-----------------|

| Sunny | Indoor | p1 | p2 |

| Sunny | Outdoor | p3 | p4 |

| Rainy | Indoor | p5 | p6 |

| Rainy | Outdoor | p7 | p8 |

In this table, p1, p2, p3, p4, p5, p6, p7, and p8 represent the conditional probabilities of winning the toss given specific combinations of "Weather" and "Venue" values. These probabilities should be estimated from the available data or domain knowledge.

The table assumes that the "Won Toss" node is conditionally independent of other variables (i.e., "Weather" and "Venue") given the class variable (match outcome). This assumption is a simplifying assumption of the Naive Bayes classifier, which assumes independence between features when making predictions