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

**Ans:**

Bayes theorem states the following:

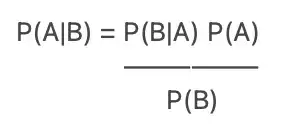
*Posterior = Prior \* Likelihood*

This can also be stated as *P (A | B) = (P (B | A) \* P(A)) / P(B)* , where *P(A|B)* is the probability of *A* given *B*, also called posterior.

**Prior**: Probability distribution representing knowledge or uncertainty of a data object prior or before observing it

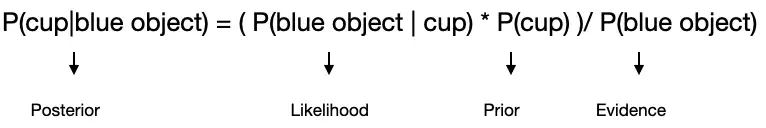
**Posterior**: Conditional probability distribution representing what parameters are likely after observing the data object

**Likelihood**: The probability of falling under a specific category or class.



P(A|B) = Posterior. P(B|A) = Likelihood. P(A) = Prior. P(B) = Evidence.

**Example:** Suppose we have 5 blue cups and 3 blue plates. We want to know if we pick any blue object between cup and plate, what the probability is if it is a cup?

So according to Bayes’ theorem probability to get a blue cup is:

P(cup|blue object) = P(blue object | cup) P(cup) / p(blue object)

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

**Ans:**

The Bayesian method of calculating conditional probabilities is used in machine learning applications that involve classification tasks. A simplified version of the Bayes Theorem, known as the Naive Bayes Classification, is used to reduce computation time and costs.

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

**Example:** Say a bookstore manager has information about his customers’ age and income. He wants to know how book sales are distributed across three age-classes of customers: youth (18-35), middle-aged (35-60), and seniors (60+).

Let us term our data X. In Bayesian terminology, X is called evidence. We have some hypothesis H, where we have some X that belongs to a certain class C.

Our goal is to determine the conditional probability of our hypothesis H given X, i.e., P(H | X).

In simple terms, by determining P(H | X), we get the probability of X belonging to class C, given X. X has attributes of age and income – let’s say, for instance, 26 years old with an income of $2000. H is our hypothesis that the customer will buy the book.

terms:

Evidence – As discussed earlier, P(X) is known as evidence. It is simply the probability that the customer will, in this case, be of age 26, earning $2000.

Prior Probability – P(H), known as the prior probability, is the simple probability of our hypothesis – namely, that the customer will buy a book. This probability will not be provided with any extra input based on age and income. Since the calculation is done with lesser information, the result is less accurate.

Posterior Probability – P(H | X) is known as the posterior probability. Here, P(H | X) is the probability of the customer buying a book (H) given X (that he is 26 years old and earns $2000).

Likelihood – P(X | H) is the likelihood probability. In this case, given that we know the customer will buy the book, the likelihood probability is the probability that the customer is of age 26 and has an income of $2000.

Given these, Bayes Theorem states:

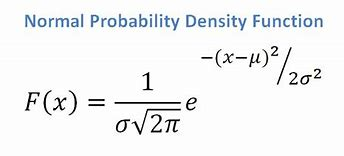
P(H | X) = [ P(X | H) \* P(H) ] / P(X)

Note the appearance of the four terms above in the theorem – posterior probability, likelihood probability, prior probability, and evidence.

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

**Ans:**

continuous attribute, considering we have a normal distribution function, we apply the following formula, with mean ? and standard deviation ?:



There are two primary ways to incorporate continuous features into the Naive Bayes model:

1. Discretization: One can transform continuous features int*o discrete features by categorizing different values into discrete buckets*. For instance, a continuous feature can be binarized by treating all values that exceed a threshold as "Large" and all values the don't "Small". Over course, more fine grained discretization that categorizes values into any arbitrary number of buckets is possible as well.
2. Distribution modeling: To incorporate continuous features into the Naive Bayes model without discretization, *one can replace the conditional probability of that feature given each class label in the Naive Bayes expression with the height of the* ***feature's conditional PDF*** *given the class label at the feature's value*. For an explanation of why this is mathematically sound in addition to intuitive, see: (How can) / (Should) I create a Naive Bayes model with different feature distributions? The heavily hand waved translation is that the probability that a continuous feature takes on a single value is 0 so we replace probabilities with PDF values as a proxy for them. Now, to calculate the height of a feature's conditional PDF given a class label at a specific feature value we need the PDF itself. A PDF can be parametrized in training, meaning its parameters can be learned from data, but its form must be predefined. The first Naive Bayes model to incorporate continuous feature values used the Gaussian distribution to model its continuous features, and was called the Gaussian Naive Bayes model as a result. However, any arbitrary probability distribution can be used to model continuous features in place of the Gaussian distribution. Commonly, data scientists use the the Poisson distribution to model features that represent random event counts, the Log-normal distribution to model non-negative quantities such as wealth distributions, and non-parametric Kernel Density Estimators to model features for which one can't predict the underlying distribution well. One can incorporate discrete features with any combination of continuous features modeled with different distributions in the same Naive Bayes model concurrently.

**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?**

**Ans:**

A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph

It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction**, and **decision making under uncertainty**

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?

Ans:

This can be solved directly with the Bayesian theorem.

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

= (P(A = 1|I = 1)P(I = 1)) / P(A = 1|I = 1)P(I = 1) + P(A = 1|I = 0)P(T = 0)

= (0.98 × 0.00001) / ((0.98 × 0.00001) +( 0.001 × (1 − 0.00001)) )

= 0.0097

≈ 0.00001/ 0.001 = 0.01

It is interesting that even though for any passenger it can be decided with high reliability (98% and 99.9%) whether (s)he is a terrorist or not, if somebody gets arrested as a terrorist, (s)he is still most likely not a terrorist (with a probability of 99%).

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).

**Ans:**

T = p means Test positive,

T = n means Test negative,

D = p means person takes drug,

D = n means person does not take drugs

We know:

P(T = p|D = n) = 0.01 (false positives)

(false negatives) P(T = n|D = p) = 0.05 =⇒ P(T = p|D = p) = 0.95 (true positives)

P(D = p) = 0.02 =⇒ P(D = n) = 0.98

We want to know the probability that somebody who tests positive is actually taking drugs:

P(D = p|T = p) = P(T = p|D = p)P(D = p) / P(T = p) (Bayes theorem)

We do not know P(T = p):

P(T = p) = P(T = p|D = p)P(D = p) + P(T = p|D = n)P(D = n) (6)

We get:

P(D = p|T = p) = P(T = p|D = p)P(D = p) / P(T = p)

=P(T = p|D = p)P(D = p) / P(T = p|D = p)P(D = p) + P(T = p|D = n)P(D = n)

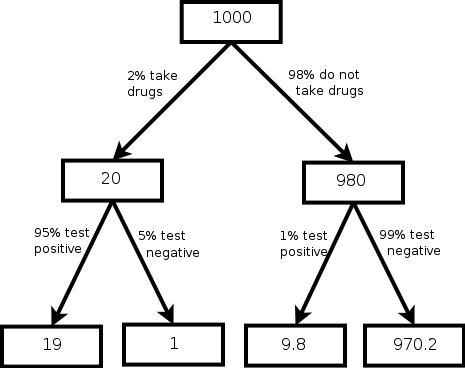
=0.95 · 0.02 / 0.95 · 0.02 + 0.01 · 0.98

= 0.019/0.0288 ≈ 0.66

There is a chance of only two thirds that someone with a positive test is actually taking drugs.

An alternative way to solve this exercise is using decision trees. Let’s assume there are 1000 people tested.

What would the result look like?



Now we can put this together in a contingency table:

|  | D = p | D = n | sum |
| --- | --- | --- | --- |
| T = p | 19 | 9.8 | 28.8 |
| T = n | 1 | 970.2 | 971.2 |
| sum | 20 | 980 | 1000 |

To determine the probability that somebody who tests positive is actually taking drugs we have to calculate:

taking drugs and positive test / all positive test = 19/28.8

≈ 0.66 (11)

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?

Ans:

The probability to solve the problem of the exam is the probability of getting a problem of

a certain type times the probability of solving such a problem, summed over all types. This is known

as the total probability.

P(solved) = P(solved|A).P(A) + P(solved|B).P(B) + P(solved|C).P(C)

= 9/10 . 30% + 2/10 . 20% + 6/10 . 50%

= 27/100 + 4/100 + 30/100

= 61/100 = 0:61 :

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

Ans: For this to answer we need Bayes theorem.

P(A|solved) = P(solved|A).P(A) / P(solved)

=(9/10 . 30%)/ 61/100

= 27/100 / 61/100 = 27/61

=0.422

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 the likelihood that there is a customer if there is a photograph?

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.