

# Assignment 13 Solutions

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

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.

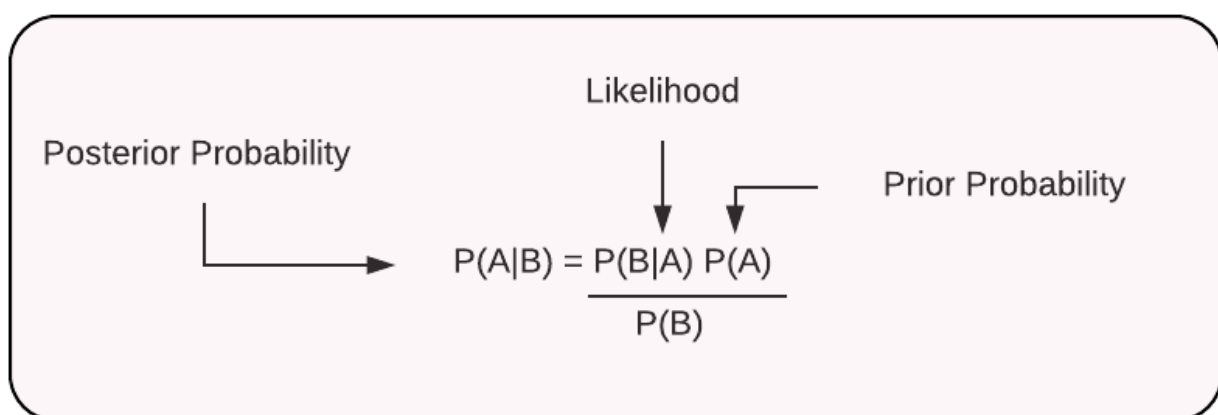
Prior: Probability distribution representing knowledge or uncertainty of a data object prior or before observing it

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.

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. Bayesian inference has found application in various activities, including medicine, science, philosophy, engineering, sports, law, etc. For example, we can use Bayes' theorem to define the accuracy of medical test results by considering how likely any given person is to have a disease and the test's overall accuracy. Bayes' theorem relies on consolidating prior probability distributions to generate posterior probabilities. In Bayesian statistical inference, prior probability is the probability of an event before new data is collected.

"



"

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

The role Bayes theorem play in the concept learninf principle is described below:

Bayes theorem helps us to calculate the single term  $P(B|A)$  in terms of  $P(A|B)$ ,  $P(B)$ , and  $P(A)$ . This rule is very helpful in such scenarios where we have a good probability of  $P(A|B)$ ,  $P(B)$ , and  $P(A)$  and need to determine the fourth term.

Naïve Bayes classifier is one of the simplest applications of Bayes theorem which is used in classification algorithms to isolate data as per accuracy, speed and classes.

Let's understand the use of Bayes theorem in machine learning with below example.

Suppose, we have a vector  $A$  with  $I$  attributes. It means

$A = A_1, A_2, A_3, A_4, \dots, A_i$

Further, we have  $n$  classes represented as  $C_1, C_2, C_3, C_4, \dots, C_n$ .

These are two conditions given to us, and our classifier that works on Machine Language has to predict  $A$  and the first thing that our classifier has to choose will be the best possible class. So, with the help of Bayes theorem, we can write it as:

$$P(C_i/A) = [P(A/C_i) * P(C_i)] / P(A)$$

Here;

$P(A)$  is the condition-independent entity.

$P(A)$  will remain constant throughout the class means it does not change its value with respect to change in class. To maximize the  $P(C_i/A)$ , we have to maximize the value of term  $P(A/C_i) * P(C_i)$ .

With  $n$  number classes on the probability list let's assume that the possibility of any class being the right answer is equally likely. Considering this factor, we can say that:

$$P(C_1) = P(C_2) = P(C_3) = P(C_4) = \dots = P(C_n).$$

This process helps us to reduce the computation cost as well as time. This is how Bayes theorem plays a significant role in Machine Learning and Naïve Bayes theorem has simplified the conditional probability tasks without affecting the precision. Hence, we can conclude that:

$$P(A_i/C) = P(A_1/C) * P(A_2/C) * P(A_3/C) * \dots * P(A_n/C)$$

Hence, by using Bayes theorem in Machine Learning we can easily describe the possibilities of smaller events.

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

**Some Real life Applications of Naive Bayes Classifier are mentioned below:**

1. Real time Prediction: Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.
2. Multi class Prediction: This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
3. Text classification/ Spam Filtering/ Sentiment Analysis: Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)

4. Recommendation System: Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not.

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

**There are two ways to estimate the class-conditional probabilities for continuous attributes in naive Bayes classifiers:**

1. We can discretize each continuous attribute and then replace the continuous attribute value with its corresponding discrete interval. This approach transforms the continuous attributes into ordinal attributes. The conditional probability  $P(X|Y=y)$ , where  $Y$  is the target variable is estimated by computing the fraction of training records belonging to class  $y$  that falls within the corresponding interval for  $X$ .
2. We can assume a certain form of probability distribution for the continuous variable and estimate the parameters of the distribution using the training data. A Gaussian distribution is usually chosen to represent the class-conditional probability for continuous attributes. The distribution is characterized by two parameters, its mean and variance.

"

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{(x-\mu_a)^2}{2\sigma_a^2}}$$

where  $\mu_a$  is the *sample mean*:

$$\mu_a = \frac{1}{|D_a|} \sum_{x \in D_a} x.a$$

$\sigma_a$  is the *sample standard deviation*, and

$\sigma_a^2$  the *sample variance*:

$$\sigma_a^2 = \frac{1}{|D_a|-1} \sum_{x \in D_a} (x.a - \mu_a)^2$$

"

#### 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 ?

**Bayesian belief network** is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty. We can define a Bayesian network as:

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

It is also called a Bayes network, belief network, decision network, or Bayesian model.

Bayesian networks are probabilistic, because these networks are built from a probability distribution, and also use probability theory for prediction and anomaly detection.

Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network. It can also be used in various tasks including prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction, and decision making under uncertainty.

**Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:**

- Directed Acyclic Graph
- Table of conditional probabilities.

The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an Influence diagram.

A Bayesian network graph is made up of nodes and Arcs (directed links), where: Each node corresponds to the random variables, and a variable can be continuous or discrete. Arc or directed arrows represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph. These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other.

**The Bayesian network has mainly two components:**

- Causal Component
- Actual numbers

Each node in the Bayesian network has condition probability distribution  $P(X_i | \text{Parent}(X_i))$ , which determines the effect of the parent on that node.

Bayesian network is based on Joint probability distribution and conditional probability. So let's first understand the joint probability distribution:

**The semantics of Bayesian Network:**

There are two ways to understand the semantics of the Bayesian network, which is given below:

1. To understand the network as the representation of the Joint probability distribution.

It is helpful to understand how to construct the network.

2. To understand the network as an encoding of a collection of conditional independence statements.

It is helpful in designing inference procedure.

**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 ?**

This can be solved directly with the Bayesian theorem.  $P(T = 1 | A = 1) = [P(A = 1 | T = 1)P(T = 1)] / P(A = 1)$

$$= P(A = 1 | T = 1)P(T = 1) / [P(A = 1 | T = 1)P(T = 1) + P(A = 1 | T = 0)P(T = 0)]$$

$$= 0.98 \times 0.00001 / [0.98 \times 0.00001 + 0.001 \times (1 - 0.00001)] = 0.0097$$

$$\approx 0.00001 / 0.001 = 0.01 \text{ Answer}$$

It is important to note 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

**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) ?**

We know:  $P(T = p|D = n) = 0.01$  (false positives) (false negatives)  $P(T = n|D = p) = 0.05 \Rightarrow P(T = p|D = p) = 0.95$  (true positives)

$P(D = p) = 0.02 \Rightarrow P(D = n) = 0.98$

We want to know the probability that somebody who tests positive is actually taking drugs:  $P(D = p|T = p) = \frac{P(T = p|D = p)P(D = p)}{P(T = p)}$  (Bayes theorem) We do not know  $P(T = p)$ :  $P(T = p) = P(T = p|D = p)P(D = p) + P(T = p|D = n)P(D = n)$

$P(D = p|T = p) = \frac{P(T = p|D = p)P(D = p)}{P(T = p|D = p)P(D = p) + P(T = p|D = n)P(D = n)} = \frac{0.95 \cdot 0.02}{0.95 \cdot 0.02 + 0.01 \cdot 0.98}$

$= 0.019/0.0288 \approx 0.66$  Answer

here is a chance of only two thirds that someone with a positive test is actually taking it.

**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?

What is the probability that you will solve the problem of the exam? Solution: 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.  $P(\text{solved}) = P(\text{solved}|A)P(A) + P(\text{solved}|B)P(B) + P(\text{solved}|C)P(C)$  (1)  $= \frac{9}{10} \cdot 30\% + \frac{2}{10} \cdot 20\% + \frac{6}{10} \cdot 50\%$  (2)  $= \frac{27}{100} + \frac{4}{100} + \frac{30}{100} = \frac{61}{100} = 0.61$  . (3) (b) Given you have solved the problem, what is the probability that it was of type A? Solution: For this to answer we need Bayes theorem.  $P(A|\text{solved}) = \frac{P(\text{solved}|A)P(A)}{P(\text{solved})}$  (4)  $= \frac{\frac{9}{10} \cdot 30\%}{\frac{61}{100}} = \frac{27/100}{61/100} = \frac{27}{61} \approx 0.442..$

**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?

Solution:

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

There are  $10 \times 12 = 120$  five-minute periods per day. In each period there is a probability of 5%. Thus the average number of customers is  $120 \times 5\% = 120 \times 0.05 = 6$

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? On average there is no photograph in  $120 - 6$  of the five-minute periods. This times the probability of 10% per period for a false photo yields  $(120 - 6) \times 10\% = 11.4$  false alarms. On average there are 6 photos, each of which has a probability of 1% of getting missed. Thus the number of false photos is  $6 \times 1\% = 0.06$
3. Explain likelihood that there is a customer if there is a photograph?

$0.99 \cdot 0.05 / [0.99 \cdot 0.05 + 0.1 \cdot (1 - 0.05)] = 0.342$

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.?

Won Toss	Match Won	Probability
Heads	Win	0.6
Heads	Lose	0.4
Tails	Win	0.3
Tails	Lose	0.7

In [ ]:

1	
---	--