BAYES THEOREM

1. **State and prove Bayes Theorem. Explain all the intermediate steps and theorems.**

Ans: - Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. Bayes theorem is also known as the formula for the probability of "causes". For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability.

Let $E_1, E_2, \ldots, E_n$ be a set of events associated with a sample space S, where all the events $E_1, E_2, \ldots, E_n$ have nonzero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes theorem,

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{k=1}^{n} P(E_k)P(A|E_k)}$$

for any k = 1, 2, 3, ..., n

If A and B are two events, then the formula for the Bayes theorem is given by:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad where \;\; P(B) \neq 0$$

Bayes Theorem Proof

According to the conditional probability formula,

$$P(E_i \mid A) = \frac{P(E_i \cap A)}{P(A)} \ldots (1)$$

Using the multiplication rule of probability,

$$P(E_i \cap A) = P(E_i)P(A \mid E_i) \ldots (2)$$

Using total probability theorem,

$$P(A) = \sum_{k=1}^{n} P(E_k)P(A|E_k) \ldots (3)$$

Putting the values from equations (2) and (3) in equation 1, we get

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{k=1}^{n} P(E_k)P(A|E_k)}$$

Note:

The following terminologies are also used when the Bayes theorem is applied:

Hypotheses: The events $E_1, E_2, \ldots E_n$ is called the hypotheses

Priori Probability: The probability $P(E_i)$ is considered as the priori probability of hypothesis E_i

Posteriori Probability: The probability $P(E_i|A)$ is considered as the posteriori probability of hypothesis E_i

Bayes' theorem is also called the formula for the probability of "causes". Since the E_i's are a partition of the sample space S, one and only one of the events E_i occurs (i.e. one of the events E_i must occur and the only one can occur). Hence, the above formula gives us the probability of a particular E_i (i.e. a "Cause"), given that the event A has occurred.

2. **Explain the significance of Bayes theorem and for what type of problems it is helpful.**

Ans: - The significance of Bayes theorem:

- It underlies many probabilistic reasoning systems in Artificial Intelligence.
- It provides an optimal way to update the probability of hypotheses given data
- It provides a principled and natural way in the estimation of the probability
- It is very much flexible and priod data which are used is realistic
- It focuses on to recover the whole area of the problems rather than the estimation of confidential estimation
- It follows the principle of likelihood which is helpful for the estimation of probability

Bayes theorem is used to find the reverse probabilities if we know the conditional probability of an event. It allows to update the predicted probabilities of an event by incorporating new information Following are some applications of Bayes' theorem:

- It is used to calculate the next step of the robot when the already executed step is given.
- It is helpful in weather forecasting.
- It can solve the Monty Hall problem.

3. **Explain Naive Bayes classifier and training process.**

Ans: - Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which help in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, which can be described as:

Naive: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an

apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

Working of Naïve Bayes' Classifier can be understood with the following steps:

- Convert the given dataset into frequency tables.
- Generate Likelihood table by finding the probabilities of given features.
- Use Bayes theorem to calculate the posterior probability.

Steps to implement Naïve Bayes algorithm using Python:

- Data Pre-processing step
- Fitting Naive Bayes to the Training set
- Predicting the test result
- Test accuracy of the result(Creation of Confusion matrix)
- Visualizing the test set result.

4. Explain Email spam filtering and how Bayes theorem can be helpful.

Ans: - Email spam filtering protects email and employees from spam that can deliver malware, zero-day attacks and other advanced threats, and spread throughout the enterprise. Incorporating an email spam filter not only helps to protect against advanced cyberattacks, but will prevent inboxes from filling up with unproductive junk mail. It's critical to use spam filtering to protect organizations.

Naive Bayes spam filtering is a baseline technique for dealing with spam that can tailor itself to the email needs of individual users and give low false positive spam detection rates that are generally acceptable to users.

Particular words have particular probabilities of occurring in spam email and in legitimate email. The filter doesn't know these probabilities in advance, and must first be trained so it can build them up. To train the filter, the user must manually indicate whether a new email is spam or not. For all words in each training email, the filter will adjust the probabilities that each word will appear in spam or legitimate email in its database.

After training, the word probabilities (also known as likelihood functions) are used to compute the probability that an email with a particular set of words in it belongs to either category. Each word in the email contributes to the email's spam probability, or only the

most interesting words. This contribution is called the posterior probability and is computed using Bayes' theorem. Then, the email's spam probability is computed over all words in the email, and if the total exceeds a certain threshold (say 95%), the filter will mark the email as a spam. As in any other spam filtering technique, email marked as spam can then be automatically moved to a "Junk" email folder, or even deleted

Bayesian email filters utilize Bayes' theorem.

- compute the probability that the message is spam, knowing that a given word appears in this message
- compute the probability that the message is spam, taking into consideration all of its words (or a relevant subset of them)
- deal with rare words