1. What is prior probability? Give an example.

A1. Prior probability is the initial degree of belief in a hypothesis before considering the evidence or data. It represents the probability of an event before new evidence or data is taken into account.

For example, suppose a doctor wants to determine the probability that a patient has a particular disease based on their symptoms. The doctor has seen many patients with similar symptoms before and knows that 10% of patients with these symptoms have the disease (prior probability). However, after conducting some tests, the doctor obtains more information about the patient's condition and updates the probability (posterior probability) based on the test results.

2. What is posterior probability? Give an example.

A2.   
Posterior probability refers to the probability of an event or hypothesis after considering new evidence or data. It is updated using Bayes' theorem, which combines the prior probability with the likelihood of observing the data to obtain the new probability.

For example, let's say a company is trying to estimate the probability of a customer buying a product based on their age. They have prior knowledge that customers above the age of 30 are more likely to make a purchase. They then collect data on a new customer who is 35 years old and observe that they have made a purchase. Using Bayes' theorem, the posterior probability of this customer making a purchase given their age can be calculated by combining the prior probability with the likelihood of observing the data. This updated probability can be used to make predictions about future customers in the same age group.

3. What is likelihood probability? Give an example.

A3. In probability theory, the likelihood function is a function of the parameters of a statistical model. It indicates how likely the observed data is given the values of those parameters. The likelihood function is often used to estimate the parameters of a model by maximizing the probability of observing the given data.

For example, consider a scenario where a pharmaceutical company is testing a new drug for a certain disease. They conduct a clinical trial on a group of patients, and record whether or not the drug was effective in each case. The likelihood function in this case would be the probability of observing the data (i.e. the number of successful treatments and unsuccessful treatments) given the parameters of the model (i.e. the efficacy of the drug). The pharmaceutical company can then use the likelihood function to estimate the efficacy of the drug and make decisions about its future use.

4. What is Naïve Bayes classifier? Why is it named so?

A4. Naïve Bayes classifier is a probabilistic machine learning algorithm used for classification tasks. It is based on the Bayes theorem, which states that the probability of a hypothesis (or event) given the evidence is proportional to the probability of the evidence given the hypothesis multiplied by the prior probability of the hypothesis.

Naïve Bayes classifier is named so because it makes the assumption that the features (or attributes) are conditionally independent given the class variable. This assumption is "naïve" because it is often not true in practice, as the features may be correlated. However, despite this simplifying assumption, Naïve Bayes has been found to be effective in many real-world applications, especially in text classification.

The Naïve Bayes classifier calculates the posterior probability of each class given the features, and selects the class with the highest probability as the predicted class for a given instance. The algorithm requires training data to estimate the prior and conditional probabilities, which can be obtained using maximum likelihood estimation or Bayesian estimation methods.

5. What is optimal Bayes classifier?

A5. The Optimal Bayes classifier is a classification algorithm that uses Bayes' theorem to make predictions. It calculates the probability of each class given the input features and assigns the class with the highest probability as the output. The Optimal Bayes classifier assumes that the input features are independent and identically distributed, and that the class-conditional distributions follow a specific parametric form (e.g., Gaussian distribution). It is considered optimal because it minimizes the classification error rate. However, it requires knowledge of the true class-conditional distributions, which may not always be available in practice.

6. Write any two features of Bayesian learning methods.

A6. Two features of Bayesian learning methods are:

1. Bayesian learning allows for the incorporation of prior knowledge or assumptions into the learning process, which can help improve the accuracy of the resulting model.
2. Bayesian learning provides a probabilistic framework for making predictions, which allows for the calculation of uncertainty or confidence in the model's predictions.

7. Define the concept of consistent learners.

A7. Consistent learners are machine learning models that guarantee to converge to the true underlying function that generated the data as the number of training examples grows to infinity. In other words, as the model observes more and more data, it becomes better at approximating the true function. This concept is based on the idea of minimizing the difference between the true distribution of the data and the learned distribution from the model, which is measured by some loss function. Consistency is a desirable property of machine learning algorithms, as it ensures that the learned model is close to the true function and can make accurate predictions on unseen data.

8. Write any two strengths of Bayes classifier.

A8. Here are two strengths of the Bayes classifier:

1. Bayes classifier can work well with small data sets: Unlike other machine learning algorithms that may require large amounts of data to produce accurate results, Bayes classifiers can work well with small data sets. This is because the algorithm is based on probabilities and can make predictions even when there are few instances of a particular feature.
2. Bayes classifier can handle irrelevant features: The algorithm can handle irrelevant features in the data set without impacting the overall accuracy of the model. This is because the algorithm calculates the probability of a given feature being associated with a particular class, and the presence or absence of irrelevant features does not significantly impact this calculation.

9. Write any two weaknesses of Bayes classifier.

A9. Here are two potential weaknesses of Bayes classifiers:

1. Assumption of independence: Bayes classifiers assume that all features are independent of each other, which may not always be the case in real-world datasets. If there are strong correlations between features, the Bayes classifier's accuracy may suffer.
2. Overfitting: If the training dataset is small or if the features are highly specific to the training set, then the Bayes classifier may overfit and not generalize well to new, unseen data. This can result in poor classification accuracy on new data.

10. Explain how Naïve Bayes classifier is used for

1. Text classification

In text classification, the Naïve Bayes classifier is used to classify documents into different categories, such as sports, politics, technology, etc. The algorithm works by analyzing the frequency of occurrence of each word in the document and calculating the probability of the document belonging to a particular category. For example, if a document contains the words "football," "team," "score," and "win," the algorithm will calculate the probability of the document belonging to the sports category. The document is then classified into the category with the highest probability.

2. Spam filtering

Naïve Bayes classifier is commonly used in spam filtering to classify incoming emails as spam or not spam. The algorithm works by analyzing the frequency of occurrence of specific words or phrases in the email and calculating the probability of the email being spam. For example, if an email contains the words "free," "money," and "offer," the algorithm will calculate the probability of the email being spam. If the probability is above a certain threshold, the email is classified as spam and filtered out.

3. Market sentiment analysis

In market sentiment analysis, Naïve Bayes classifier is used to analyze the sentiment of market news or social media posts about a particular stock or company. The algorithm works by analyzing the frequency of positive, negative, and neutral words in the text and calculating the probability of the sentiment being positive or negative. For example, if a news article about a company contains words like "growth," "profit," and "success," the algorithm will calculate the probability of the sentiment being positive. The sentiment is then classified as positive or negative based on the probability threshold set by the analyst. This information can be used by traders and investors to make informed decisions about buying or selling a stock.