**1. What is prior probability? Give an example.**

Prior probability shows the likelihood of an outcome in a given dataset. P(H) and P(E) are called prior probabilities. They are probabilities assigned prior to acquisition of new data and are updated with the acquisition of new data. For example, if you are classifying the buyers of a specific car, you might already know that 60% of purchasers are male and 40% are female. If you know or can estimate these probabilities, a discriminant analysis can use these prior probabilities in calculating the posterior probabilities.

**2. What is posterior probability? Give an example.**

Posterior probability is used in a wide variety of domains including finance, medicine, economics, and weather forecasting.

The whole point of using posterior probabilities is to update a previous belief we had about something once we obtain new information.

For example that the probability of a given tree in the forest being Oak was 20%. This is known as a **prior probability**. If we simply picked a tree at random, we knew that the probability of it being an Oak was 0.20.

However, once we obtained the new information that the tree we selected was healthy, we were able to use this new information to determine that the **posterior probability** of this tree being an Oak was instead 0.3103.

**3. What is likelihood probability? Give an example.**

Likelihood refers to the process of determining the best data distribution given a specific situation in the data. For example, in sentiment analysis, probability is used to determine the likelihood that a given text expresses positive, negative, or neutral sentiment.

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

Naive Bayes algorithm (NB) is Bayesian graphical model that has nodes corresponding to each of the columns or features. It is called naive because, it ignores prior distribution of parameters and assume independence of all features and all rows.

**5. What is optimal Bayes classifier?**

The Bayes Optimal Classifier is a probabilistic model that uses training data and hypotheses to make predictions for new data. It is also known as the target classifier.

**6. Write any two features of Bayesian learning methods.**

• Each observed training example can incrementally decrease or increase the estimated probability that a hypothesis is correct.   
– This provides a more flexible approach to learning than algorithms that completely eliminate a hypothesis if it is found to be inconsistent with any single example.   
• Prior knowledge can be combined with observed data to determine the final probability of a hypothesis. In Bayesian learning, prior knowledge is provided by asserting   
– a prior probability for each candidate hypothesis, and – a probability distribution over observed data for each possible hypothesis.   
• Bayesian methods can accommodate hypotheses that make probabilistic predictions   
• New instances can be classified by combining the predictions of multiple hypotheses, weighted by their probabilities.   
• Even in cases where Bayesian methods prove computationally intractable, they can provide a standard of optimal decision making against which other practical methods can be measured.

**7. Define the concept of consistent learners.**

Consistent Learners. • A learner L using a hypothesis H and training data D is said to be a consistent learner if it always outputs a hypothesis with zero error on D whenever H contains such a hypothesis. • By definition, a consistent learner must produce a hypothesis in the version space for H given D.

**8. Write any two strengths of Bayes classifier.**

* It is simple and easy to implement.
* It doesn't require as much training data.
* It handles both continuous and discrete data.
* It is highly scalable with the number of predictors and data points.
* It is fast and can be used to make real-time predictions.

**9. Write any two weaknesses of Bayes classifier.**

* it relies on an often-faulty assumption of equally important and independent features which results in biased posterior probabilities. Although this assumption is rarely met, in practice, this algorithm works surprisingly well.  
  Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.
* This algorithm faces the ‘zero-frequency problem’ where it assigns zero probability to a categorical variable whose category in the test data set wasn’t available in the training dataset. It would be best if you used a smoothing technique to overcome this issue.
* Its estimations can be wrong in some cases, so you shouldn’t take its probability outputs very seriously.

**10. Explain how Naïve Bayes classifier is used for**

**1. Text classification:** Most of the time, Naive Bayes finds uses in-text classification due to its assumption of independence and high performance in solving multi-class problems. It enjoys a high rate of success than other algorithms due to its speed and efficiency.

**2. Spam filtering:**

1. Find the probability of each word: Find the probability that each word in the email is spam.
2. Multiply the probabilities together: Multiply the probabilities together to get the overall email spam metric.
3. Use the metric for classification: Use the metric for classification.

**3. Market sentiment analysis:**

1. It assumes that features are independent given the sentiment.
2. It calculates the probability of a text belonging to each sentiment class based on word frequencies.
3. It assigns the class with the highest probability.