Machine Learning Algorithm

Interview Questions & Answers

1.What is the difference between classification and regression?

Classification is used to produce discrete results, classification is used to classify data into some specific categories. For example, classifying emails into spam and non-spam categories. Whereas, We use regression analysis when we are dealing with continuous data, for example predicting stock prices at a certain point in time.

2.What are the parametric models? Give an example.

Parametric models are those with a finite number of parameters. To predict new data, you only need to know the parameters of the model. Examples include linear regression, logistic regression, and linear SVMs. Non-parametric models are those with an unbounded number of parameters, allowing for more flexibility. To predict new data, you need to know the parameters of the model and the state of the data that has been observed. Examples include decision trees, k-nearest neighbors, and topic models using latent Dirichlet analysis.

3.Explain the difference between supervised and unsupervised machine learning?

In supervised machine learning algorithms, we have to provide labeled data, for example, prediction of stock market prices, whereas in unsupervised we need not have labeled data, for example, classification of emails into spam and non-spam.

4.What Is Overfitting, and How Can You Avoid It?

Overfitting is a situation that occurs when a model learns the training set too well, taking up random fluctuations in the training data as concepts. These impact the model's ability to generalize and don't apply to new data. When a model is given the training data, it shows 100 percent accuracy—technically a slight loss. But, when we use the test data, there may be an error and low efficiency. This condition is known as overfitting.

There are multiple ways of avoiding overfitting, such as:

● Regularization. It involves a cost term for the features involved with the objective function

● Making a simple model. With lesser variables and parameters, the variance can be reduced

● Cross-validation methods like k-folds can also be used

● If some model parameters are likely to cause overfitting, techniques for regularization like LASSO can be used that penalize these parameters

5.What is meant by 'Training set' and 'Test Set'?

We split the given data set into two different sections namely,'Training set' and 'Test Set'. 'Training set' is the portion of the dataset used to train the model. 'Testing set' is the portion of the dataset used to test the trained model.

6.How Do You Handle Missing or Corrupted Data in a Dataset?

One of the easiest ways to handle missing or corrupted data is to drop those rows or columns or replace them entirely with some other value.

There are two useful methods in Pandas:

● IsNull() and dropna() will help to find the columns/rows with missing data and drop them

● Fillna() will replace the wrong values with a placeholder value

7.Explain Ensemble learning.

In ensemble learning, many base models like classifiers and regressors are generated and combined together so that they give better results. It is used when we build component classifiers that are accurate and independent. There are sequential as well as parallel ensemble methods.

8.Explain the Bias-Variance Tradeoff.

Predictive models have a tradeoff between bias (how well the model fits the data) and variance (how much the model changes based on changes in the inputs). Simpler models are stable (low variance) but they don't get close to the truth (high bias). More complex models are more prone to overfitting (high variance) but they are expressive enough to get close to the truth (low bias). The best model for a given problem usually lies somewhere in the middle.

9.What is the difference between stochastic gradient descent (SGD) and gradient descent (GD)?

Both algorithms are methods for finding a set of parameters that minimize a loss function by evaluating parameters against data and then making adjustments. In standard gradient descent, you'll evaluate all training samples for each set of parameters. This is akin to taking big, slow steps toward the solution. In stochastic gradient descent, you'll evaluate only 1 training sample for the set of parameters before updating them. This is akin to taking small, quick steps toward the solution.

10.How Can You Choose a Classifier Based on a Training Set Data Size?

When the training set is small, a model that has a right bias and low variance seems to work better because they are less likely to overfit. For example, Naive Bayes works best when the training set is large. Models with low bias and high variance tend to perform better as they work fine with complex relationships.

11.What are 3 data preprocessing techniques to handle outliers?

1. Winsorize (cap at threshold).

2. Transform to reduce skew (using Box-Cox or similar).

3. Remove outliers if you're certain they are anomalies or measurement errors.

12. How much data should you allocate for your training, validation, and test sets?

You have to find a balance, and there's no right answer for every problem. If your test set is too small, you'll have an unreliable estimation of model performance (performance statistics will have high variance). If your training set is too small, your actual model parameters will have a high variance. A good rule of thumb is to use an 80/20 train/test split. Then, your train set can be further split into train/validation or into partitions for cross-validation.

13.What Is a False Positive and False Negative and How Are They Significant?

False positives are those cases which wrongly get classified as True but are False. False negatives are those cases which wrongly get classified as False but are True. In the term ‘False Positive,’ the word ‘Positive’ refers to the ‘Yes’ row of the predicted value in the confusion matrix. The complete term indicates that the system has predicted it as a positive, but the actual value is negative.

14.Explain the difference between L1 and L2 regularization.

L2 regularization tends to spread error among all the terms, while L1 is more binary/sparse, with many variables either being assigned a 1 or 0 in weighting. L1 corresponds to setting a Laplacean prior to the terms, while L2 corresponds to a Gaussian prior.

15.What’s a Fourier transform?

A Fourier transform is a generic method to decompose generic functions into a superposition of symmetric functions. Or as this more intuitive tutorial puts it, given a smoothie, it’s how we find the recipe. The Fourier transform finds the set of cycle speeds, amplitudes, and phases to match any time signal. A Fourier transform converts a signal from time to frequency domain — it’s a very common way to extract features from audio signals or other time series such as sensor data.

Thank You!
To know more Get In Touch!

Kick start your Data Science Career

Book Mentoring Session

Analytics Blog

Decoding Data Science