

Company Name - Pwc
Role - Data Scientist

Q. bias-variance trade-off?

A. The bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

Q. equation to calculate the precision and recall rate.?

A. Precision = True positives / (True positives + False positives) = TP / (TP + FP).
Recall = TruePositives / (TruePositives + FalseNegatives) = TP / (TP + FN).

Q. Normalisation and Standardization?

A. Normalization typically means rescales the values into a range of [0,1].
Standardization typically means rescales data to have a mean of 0 and a standard deviation of 1 (unit variance).

Q. Difference between Point Estimates and Confidence Interval?

A. The two are closely related. In fact, the point estimate is located exactly in the middle of the confidence interval. However, confidence intervals provide much more information and are preferred when making inferences.

Q. Statistical power of sensitivity?

A. The statistical power of an A/B test refers to the test's sensitivity to certain magnitudes of effect sizes. More precisely, it is the probability of observing a statistically significant result at level alpha (α) if a true effect of a certain magnitude (MEI) is in fact present.

Q. TF/IDF vectorization?

A. tf-idf vectorization gives a numerical representation of words entirely dependent on the nature and number of documents being considered. The same words will have different vector representations in another corpus.

Q. Entropy and Information gain in Decision tree algorithm?

A. Entropy can be defined as a measure of the purity of the sub split. Entropy always lies between 0 to 1. The information gain is based on the decrease in entropy after a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain (i.e., the most homogeneous branches).

Q. What is feature scaling and why is it necessary?

A. Feature scaling in machine learning is one of the most critical steps during the pre-processing of data before creating a machine learning model. Scaling can make a difference between a weak machine learning model and a better one.

The most common techniques of feature scaling are Normalization and Standardization.

Q. Difference between likelihood and probability?

A. Probability basically corresponds to finding the chance of something given a sample distribution of the data, while on the other hand, likelihood refers to finding the best distribution of the data given a particular value of some feature or some situation in the data.

While calculating the probability, feature value can be varied, but the characteristics(mean, standard and deviation) of the data distribution cannot be altered.

The likelihood in very simple terms means to increase the chances of a particular situation to happen by varying the characteristics of the dataset distribution.

Q. Difference between Sigmoid and Softmax function?

A. The Sigmoid Activation Function is a mathematical function with a recognizable “S” shaped curve. It is used for the logistic regression and basic neural network implementation. If we want to have a classifier to solve a problem with more than one right answer, the Sigmoid Function is the right choice. We should apply this function to each element of the raw output independently. The return value of Sigmoid Function is mostly in the range of values between 0 and 1 or -1 and 1.

Whereas the Softmax Activation Function, also known as SoftArgMax or Normalized Exponential Function is a fascinating activation function that takes vectors of real numbers as inputs, and normalizes them into a probability distribution proportional to the exponentials of the input numbers. Before applying, some input data could be negative or greater than 1. Also, they might not sum up to 1. After applying Softmax, each element will be in the range of 0 to 1, and the elements will add up to 1. This way, they can be interpreted as a probability distribution. For more clarification, the larger the input number, the larger the probabilities will be.

Q. Which machine learning algorithm is known as the lazy learner?

A. KNN is a Machine Learning algorithm known as a lazy learner. K-NN is a lazy learner because it doesn't learn any machine learnt values or variables from the training data but dynamically calculates distance every time it wants to classify, hence memorises the training dataset instead.

Q. Difference between Forecasting and Prediction.

A. Prediction is concerned with estimating the outcomes for unseen data. ... Forecasting is a sub-discipline of prediction in which we are making predictions about the future, on the basis of time-series data. Thus, the only difference between prediction and forecasting is that we consider the temporal dimension.

Q. What is Hypothesis testing?

A. Hypothesis testing in statistics is a way for you to test the results of a survey or experiment to see if you have meaningful results. You're basically testing whether your results are valid by figuring out the odds that your results have happened by chance. If your results may have happened by chance, the experiment won't be repeatable and so has little use.

Q. Difference between Regression and Time-series?

A. Time series is the word used to describe data which is ordered by time; example stock prices by date. Regression is the word used to describe a mathematical model which aims to check whether a variable, example, a man's weight is dependent on some other variables, example, his height and the amount of rice he eats every day. A possible example where both these words are used is when you say, "I want to do a regression on a time series data". The meaning of this is to check whether stock prices tomorrow are dependent on today's stock prices plus some other variables.

Q. Difference between null hypothesis and alternate hypothesis?

A. The null hypothesis is often an initial claim that is based on previous analyses or specialized knowledge. The alternative hypothesis states that a population parameter is smaller, greater, or different than the hypothesized value in the null hypothesis.

NOTE:

If you want to learn Data Science from scratch and become interview ready, check out our popular Data science course

<https://ai.cloudymL.com/Learn-Data-Science-from-Scratch>