Day 8 of 100 Data Science Interview Questions Series!!

Q 36.) Given a data set of features X and labels y , what assumptions are made when using Naive Bayes methods?

The Naive Bayes algorithm assumes that the features of X are conditionally independent of each other for the given Y.

The idea that each feature is independent of each other may not always be true, but we assume it to be true to apply Naive Bayes. This "naive" assumption is where the namesake comes from.

Q 37.) What is a Box-Cox Transformation?

A Box Cox transformation is a way to transform non-normal dependent variables into a normal shape. Normality is an important assumption for many statistical techniques, if your data isn't normal, applying a Box-Cox means that you are able to run a broader number of tests.

The residuals could either curve as the prediction increases or follow skewed distribution. In such scenarios, it is necessary to transform the response variable so that the data meets the required assumptions.

A Box cox transformation is a statistical technique to transform non-normal dependent variables into a normal shape. If the given data is not normal then most of the statistical techniques assume normality.

Q 38.) Where do you use TF/IDF vectorization?

The tf-idf is short for term frequency-inverse document frequency. It is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.

It is often used as a weighting factor in information retrieval and text mining. The tf-idf value increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general.

Q 39.) Tell me about Pattern Recognition and what areas in which it is used?

Follow: Alaap Dhall on LinkedIn for more!

Pattern recognition is the process of recognizing patterns by using machine learning algorithm. Pattern recognition can be defined as the classification of data based on knowledge already gained or on statistical information extracted from patterns and/or their representation.

Pattern Recognition can be used in

- Computer Vision
- Speech Recognition
- Data Mining
- Statistics
- Informal Retrieval
- Bio-Informatics

Q 40.) What is the difference between Type I vs Type II error?

A type I error occurs when the null hypothesis (H0) is true but is rejected. It is asserting something that is absent, a false hit. A type I error may be likened to a so-called false positive (a result that indicates that a given condition is present when it actually is not present).

A type II error occurs when the null hypothesis is false, but erroneously fails to be rejected. It is failing to assert what is present, a miss.
A type II error may be compared with a so-called false negative (where an actual 'hit' was disregarded by the test and seen as a 'miss') in a test checking for a single condition with a definitive result of true or false.

Table of error types		Null hypothesis (H_0) is	
		True	**False**
Decision About Null Hypothesis (H_0)	**Reject**	Type I error (False Positive)	Correct inference (True Positive)
	Fail to reject	Correct inference (True Negative)	Type II error (False Negative)

- Alaap Dhall

Follow: Alaap Dhall on LinkedIn for more!

Follow Alaap Dhall on LinkedIn for more insights in Data Science and Deep Learning!!

Visit https://www.aiunquote.com for a 100 project series in Deep Learning.