Day 6 of 100 Data Science Interview Questions Series!!

Q 26.) How can you use eigenvalue or eigenvector?

It is difficult to understand and visualize data with more than 3 dimensions, let alone a dataset of over 100+ dimensions. Hence, it would be ideal to somehow compress/transform this data into a smaller dataset. This is where we can use this concept.

We can utilise Eigenvalues and Eigenvectors to reduce the dimension space ensuring most of the key information is maintained.

Eigenvalues are the directions along which a particular linear transformation acts by flipping, compressing, or stretching.

Eigenvectors are for understanding linear transformations. In data analysis, we usually calculate the eigenvectors for a correlation or covariance matrix.

Please view this article which has explained this concept better than I ever could!
https://medium.com/fintechexplained/what-are-eigenvalues-and-eigenvectors-a-must-know-concept-for-machine-learning-80d0fd330e47

Q 27.) What is lemmatization and Stemming, Which one should I use in Sentimental Analysis, and which one should I use in QnA bot?

They are used as Text Normalization techniques in NLP for preprocessing text.

Stemming is the process of reducing inflection in words to their root forms such as mapping a group of words to the same stem even if the stem itself is not a valid word in the Language."

Lemmatization, unlike Stemming, reduces the inflected words properly ensuring that the root word belongs to the language. In Lemmatization root word is called Lemma.

- Stemming is better option for Sentimental Analysis as the meaning of the word is not necessary for understanding sentiments, and stemming is little faster than Lemmatization.
- Lemmatization is better for QnA bot as word should have a proper emaning while conversing with a human subject.

Q 28.) What are some common Recommendation System Types, where can I use them?

Recommendation systems are used to recommend or generate some outputs based on previous inputs that were given by users. Recommendation system can be build thorugh Deep Learning, like Deep Beliefe networks, RBM,AutoEncoder etc or some traditional techniques.

Some common types are:

1. Collaborative Recommender system
2. Content-based recommender system
3. Demographic based recommender system
4. Utility based recommender system
5. Knowledge based recommender system
6. Hybrid recommender system.

- DL based Recommendation systems can be used for dimentionality reduction and generating similar output.
- RS can also be used for suggestions of similar items based on user's past choices and item's content.
- RS can also be used for suggestions of simlar products based on a group of users with simlar features as you.

Q 29.) What is bias, variance trade-off?

Bias is error introduced in your model due to over simplification of machine learning algorithm." It can lead to under fitting.

- Low bias machine learning algorithms — Decision Trees, k-NN and SVM
- High bias machine learning algorithms — Linear Regression, Logistic Regression

Variance is error introduced in your model due to complex machine learning algorithm, your model learns noise also from the training data set and performs bad on test data set. It can lead high sensitivity and over fitting.

Normally, as you increase the complexity of your model, you will see a reduction in error due to lower bias in the model. However, this only happens till a particular point. **As you continue to make your model more complex, you end up over-fitting your model and hence your model will start suffering from high variance. Increasing the bias will decrease the variance. Increasing the variance will decrease the bias.** This is Bias-Variance Trade-Off.

Q 30.) What are vanishing/exploding gradients?

Gradient is the direction and magnitude calculated during training of a neural network that is used to update the network weights in the right direction and by the right amount.

- Exploding gradient is a problem where large error gradients accumulate and result in very large updates to neural network model weights during training.
- Vanishing gradient is a problem where as more layers are added to neural networks, the gradients of the loss function approaches zero, making the network hard to train. This occurs in large models with many layers. **Models like ResNet, that have skip connections, are a good solution to this problem.**

\- Alaap Dhall

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.