Top 20 CBOW (Continuous Bag of Words) Interview Questions with Answers

1. What is CBOW in Word2Vec?

CBOW is a Word2Vec model that predicts a target word based on its context (surrounding words) using a shallow neural network.

2. How does CBOW differ from Skip-gram?

CBOW predicts the center word from context, while Skip-gram predicts surrounding context from the center word.

3. What does it mean when CBOW is called a shallow neural network?

It has only one hidden layer (embedding layer) between input and output.

4. Why do we average the context word embeddings in CBOW?

To get a single vector representing all context words, making prediction easier and faster.

5. What is the objective function of CBOW?

To maximize the probability of predicting the correct target word given context.

6. What are the input and output of the CBOW model?

Input: context words; Output: probability distribution over vocabulary for the target word.

7. What is the role of softmax in CBOW?

It converts raw scores into probabilities over the vocabulary.

8. What kind of data is needed to train a CBOW model?

Text corpus with generated context-target pairs using a sliding window.

9. What are the dimensions of the embedding and output matrices in CBOW?

Embedding: [vocab_size x embedding_dim], Output: [embedding_dim x vocab_size].

10. Why is one-hot encoding used in CBOW?

To uniquely represent each word and fetch corresponding embeddings.

11. How is the loss calculated in CBOW?

Using cross-entropy between predicted softmax output and actual target word.

12. What is cross-entropy loss and why is it used in CBOW?

It measures the difference between predicted and true distributions, penalizing wrong predictions.

13. How does CBOW learn word embeddings during training?

By backpropagating the loss and updating embedding weights to reduce prediction error.

Top 20 CBOW (Continuous Bag of Words) Interview Questions with Answers

14. What is the gradient update step in CBOW?

Weights are updated using gradient descent: weight -= learning_rate * gradient.

15. What does the embedding matrix represent in CBOW?

Dense vector representations of words that capture meaning.

16. When would you prefer CBOW over Skip-gram?

When training on large data and focusing on frequent words.

17. What are the advantages and disadvantages of CBOW?

Advantages: Fast, simple; Disadvantages: Not good with rare words or word order.

18. Can CBOW handle out-of-vocabulary (OOV) words? How?

Not directly. Solutions include using <UNK> tokens or models like FastText.

19. How does CBOW capture semantic relationships like 'king - man + woman = queen'?

Through similar vector directions in embedding space that encode semantic meaning.

20. How do you use pre-trained CBOW embeddings in NLP tasks?

Load the embeddings and use them as input features in models for tasks like classification or NER.