1. What is the objective of the word2vec model and how does it achieve it?
2. Explain the difference between the CBOW and Skip-gram architectures for the word2vec model.
3. How does negative sampling help in training the word2vec model?
4. Describe the role of the softmax function in the training process of the word2vec model.
5. Explain the concept of subsampling in the word2vec model and how it helps in training.
6. How does the size of the training window affect the quality of the learned word embeddings in the word2vec model?
7. What is the effect of the number of negative samples on the quality of the learned word embeddings in the word2vec model?
8. How can we evaluate the quality of the learned word embeddings in the word2vec model?
9. How can we use the word2vec model for tasks such as word similarity and analogy solving?
10. Can the word2vec model be used for languages other than English? What modifications may be needed for this purpose?
11. Explain the concept of hierarchical softmax in the word2vec model and how it improves the efficiency of training.
12. The word2vec model assumes that each word has only one vector representation. Discuss the limitations of this assumption and explain how polysemous words can be handled in the model.
13. In the skip-gram model of the word2vec, the probability of predicting the context words given the center word is maximized. Suggest modifications to the model objective to take into account the order of the context words and their frequency.
14. The word2vec model learns word embeddings based on the context of the words in the training data. Discuss the limitations of this approach in capturing semantic and syntactic relations between words.
15. Explain the concept of dynamic contexts in the word2vec model and how it can be used to capture the evolving meanings of words over time.
16. The word2vec model requires a large amount of training data to learn accurate word embeddings. Suggest methods to improve the quality of the embeddings when only a limited amount of training data is available.
17. The word2vec model can be used to learn embeddings for phrases and sentences as well. Explain how this can be achieved using techniques such as the paragraph vector model and the skip-thought model.
18. Explain the concept of multi-task learning in the context of word2vec and how it can be used to improve the quality of word embeddings for downstream tasks.
19. The word2vec model can be extended to handle other types of input such as images and audio. Discuss the challenges and opportunities of applying the word2vec model to these domains.
20. The word2vec model is trained using stochastic gradient descent (SGD). Discuss the limitations of SGD in optimizing complex non-convex objectives and suggest alternative optimization methods that can be used for the word2vec model.
21. Suppose you have a vocabulary of 100 words and are using a skip-gram model with a word vector dimensionality of 50. If you use negative sampling with 5 negative samples, how many parameters does your model have?
22. Consider a corpus of 1 million sentences and a vocabulary of 50,000 words. You want to train a word2vec model using stochastic gradient descent with a batch size of 128 and a learning rate of 0.01. If each sentence contains an average of 15 words, how many epochs will it take to complete one pass over the entire corpus?
23. Suppose you are training a word2vec model using the CBOW architecture and a vocabulary of 10,000 words. You use a softmax loss function and perform gradient descent with a learning rate of 0.1. After training for 1000 iterations, you find that the loss has converged to 2.5. If you increase the learning rate to 0.2, how many iterations will it take to converge to a loss of 2.5 again?
24. Consider a word2vec model with a vocabulary of 20,000 words and a word vector dimensionality of 300. Suppose you want to cluster the word vectors into 100 clusters using k-means. How many iterations of the k-means algorithm will it take to converge to a stable clustering, assuming a convergence threshold of 0.01?