NLP assignments-6

1. Vanilla autoencoders are a type of neural network that are trained to reconstruct their input data. They are composed of an encoder and a decoder, which are neural networks that respectively transform the input data into a lower-dimensional representation (latent space) and reconstruct the original data from the latent space.
2. Sparse autoencoders are a variant of vanilla autoencoders that are trained to enforce sparsity on the activations of the hidden units in the encoder. This means that, during training, the autoencoder is encouraged to use as few hidden units as possible to represent the input data.
3. Denoising autoencoders are a variant of vanilla autoencoders that are trained to reconstruct the original input data from a corrupted version of it. They are typically used to learn robust representations of the input data that are able to withstand the presence of noise or corruption.
4. Convolutional autoencoders are a variant of vanilla autoencoders that are designed to process data with a grid-like topology, such as images. They use convolutional layers in the encoder and decoder, which allow them to take advantage of the local correlations present in the input data.
5. Stacked autoencoders are a type of deep neural network that is composed of multiple layers of autoencoders stacked on top of each other. They can be trained using an unsupervised learning algorithm, and can then be fine-tuned using supervised learning for a specific task.
6. To generate sentences using LSTM autoencoders, you can train an autoencoder with an LSTM encoder and an LSTM decoder on a dataset of sentences. The encoder will learn to compress the input sentences into a lower-dimensional representation, and the decoder will learn to reconstruct the original sentences from the latent space. Once the autoencoder is trained, you can use the decoder to generate new sentences by sampling from the latent space and feeding the samples into the decoder.
7. Extractive summarization is a method of generating a summary of a document by selecting a subset of the most important sentences or phrases from the original document.
8. Abstractive summarization is a method of generating a summary of a document by generating new phrases and sentences that capture the main ideas of the original document.
9. Beam search is a search algorithm that is often used in natural language processing to find the most likely sequence of words in a language model. It works by maintaining a set of partial hypotheses (called the "beam"), and expanding each hypothesis by adding the most likely next word at each step.
10. Length normalization is a technique used to adjust the scores of hypotheses in a beam search algorithm based on their length. This can be useful in natural language processing to prevent the beam from being dominated by very short or very long hypotheses.
11. Coverage normalization is a technique used to adjust the scores of hypotheses in a beam search algorithm based on the amount of the input document that they cover. This can be useful in natural language processing to encourage the beam to generate summaries that cover a larger portion of the input document.
12. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a metric used to evaluate the quality of summaries by comparing them to a reference summary. There are several variants of the ROUGE metric, including ROUGE-N, which compares the n-gram overlap between the summary and the reference, and ROUGE-L, which compares the longest common subsequence between the summary and the reference.

Top of Form