**NLP Assignment 6 Solution**

**1. What are Vanilla autoencoders**

**Ans:** Vanilla autoencoders are a type of neural network architecture that uses an encoder-decoder model to learn a compressed representation of data, known as an encoding, and then reconstructs the data from the encoding. The encoder part of the network takes in the input data and compresses it into a smaller representation, while the decoder part takes this smaller representation and reconstructs the original data as accurately as possible. The goal of an autoencoder is to learn a representation that captures the important features of the input data while minimizing the size of the representation.

**2. What are Sparse autoencoders**

**Ans:** Sparse autoencoders are a type of artificial neural network which uses a sparse data representation as a way of reducing the amount of data used in the network. The sparse representation is achieved by a regularization technique called sparsity, which forces the encoder to only use a small fraction of the available input neurons. This technique is used for feature learning and has been successfully applied to image recognition, text understanding, and deep learning.

**3. What are Denoising autoencoders**

**Ans:** Denoising autoencoders are a type of autoencoder neural network used for unsupervised learning. They are used to reduce noise from a signal by learning a representation of the input data. Denoising autoencoders work by adding noise to the input data, then training the network to reconstruct the original input from the noisy version. This helps the network to learn a robust representation of the data, which can be used for tasks such as classification and clustering.

**4. What are Convolutional autoencoders**

**Ans:** Convolutional autoencoders are a type of autoencoder that use convolutional layers in the encoding and decoding of the data. They are used to learn useful features from the data and can be used for tasks such as image classification, image segmentation, and image generation.

**5. What are Stacked autoencoders**

**Ans:** Stacked autoencoders are a type of deep learning neural network composed of multiple layers of autoencoders. These autoencoders are arranged in a stack, with each layer receiving input from the layer below it. The stacked autoencoders are used for feature extraction and representation learning. Each layer of the stack learns a representation of the data, which is then used as input to the next layer. This way, the deep learning model can learn a hierarchy of features and representations, allowing it to better capture the underlying structure of the data. Stacked autoencoders are a powerful tool for unsupervised learning, and have been used for a variety of tasks such as image recognition, dimensionality reduction, and anomaly detection.

**6. Explain how to generate sentences using LSTM autoencoders**

**Ans:** LSTM autoencoders are a type of recurrent neural network that is trained to generate text. The autoencoder is trained to take in a sequence of words and generate a corresponding output sequence. It does this by learning to map the input sequence to an internal representation and then reconstructing the output sequence from that representation. The autoencoder can then be used to generate new sentences from a given input by using the internal representation to generate new words or phrases. To generate sentences using an LSTM autoencoder, the model must first be trained on a large corpus of text, such as a book or a collection of news articles. Once trained, the autoencoder can be used to generate new sentences by providing it with a seed sentence or phrase. The autoencoder will then generate a sequence of words based on the seed phrase. This process can be repeated until a satisfactory sentence or paragraph is generated.

**7. Explain Extractive summarization**

**Ans:** Extractive summarization is the process of automatically creating a summary by identifying and extracting relevant phrases and sentences from a source document. It involves selecting important pieces of text from the source document and concatenating them to form a summary. It is a type of text summarization technique that focuses on finding the most important sentences or phrases in a document and extracting them to create a summary. This type of summarization is useful for extracting key facts and ideas from a large body of text, such as a book or article, and condensing them into a shorter summary.

**8. Explain Abstractive summarization**

**Ans:** Abstractive summarization is a type of summarization technique that generates new phrases and sentences to accurately capture the meaning and essence of the original text. Unlike extractive summarization, abstractive techniques rephrase the original text and may include the author's own words and interpretations. Abstractive summarization methods use natural language processing and deep learning algorithms to generate summaries. They are often used in text summarization, question answering, document summarization, and machine translation tasks.

**9. Explain Beam search**

**Ans:** Beam search is an algorithm used in artificial intelligence to find the most probable sequence of words in natural language processing, machine translation, and other applications. It is an extension of the breadth-first search algorithm. The main difference between beam search and breadth-first search is that beam search uses a fixed-width beam instead of exploring all nodes at a given depth. This helps to reduce the amount of time required to find a solution, while still allowing the algorithm to explore multiple paths. Beam search also uses heuristics, such as the number of words in the sentence, to prune the search space and make the algorithm more efficient.

**10. Explain Length normalization**

**Ans:** Length normalization is a technique used to normalize the lengths of text strings in a corpus in order to make them comparable. The goal is to make sure that the same text strings are being compared, regardless of their length. This is often done by dividing the length of a text string by the maximum length of a text string in the corpus. This helps to ensure that all text strings are being compared on an equal basis, regardless of their length.

**11. Explain Coverage normalization**

**Ans:** Coverage normalization is a method used to normalize the coverage of sequencing reads across multiple samples. It is used to account for the uneven coverage of sequencing reads due to varying library sizes or sequencing depth across samples. By normalizing coverage, researchers can compare samples with different sequencing depths on a more equal footing. Coverage normalization is often accomplished by calculating a “scaling factor” for each sample, which is used to multiply the sequencing depth of that sample to match the median sequencing depth of the entire dataset. This then allows the comparison of sequencing data from different samples on a normalized basis.

**12. Explain ROUGE metric evaluation**

**Ans:** ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a metric evaluation used to measure the quality of a summarization task. It is used to compare the automatically produced summary with a set of reference summaries that were created by humans. The ROUGE evaluation compares the summaries of two different systems and computes a score based on the number of overlapping n-grams between them. It also takes into account the length of the summary by comparing it to the average length of the reference summaries. ROUGE is a popular metric for measuring summarization quality and is used in many text summarization tasks.