**NLP- ASSIGNMENT-6**

1. **What are Vanilla autoencoders.**

The vanilla autoencoder, as proposed by Hinton, consists of only one hidden layer. The number of neurons in the hidden layer is less than the number of neurons in the input (or output) layer. This results in producing a bottleneck effect on the flow of information in the network, and therefore we can think of the hidden layer as a bottleneck layer, restricting the information that would be stored. Learning in the autoencoder consists of developing a compact representation of the input signal at the hidden layer so that the output layer can faithfully reproduce the original input.

1. **What are Sparse autoencoders.**

A Sparse Autoencoder is a type of autoencoder that employs sparsity to achieve an information bottleneck. Specifically the loss function is constructed so that activations are penalized within a layer.

1. **What are Denoising autoencoders.**

A Denoising Autoencoder is a modification on the autoencoder to prevent the network learning the identity function. Specifically, if the autoencoder is too big, then it can just learn the data, so the output equals the input, and does not perform any useful representation learning or dimensionality reduction.

1. **What are Convolutional autoencoders**

Convolutional Autoencoder is a variant of Convolutional Neural Networks that are used as the tools for unsupervised learning of convolution filters. They are generally applied in the task of image reconstruction to minimize reconstruction errors by learning the optimal filters.

1. **What are Stacked autoencoders**

A stacked denoising autoencoder is simply many denoising autoencoders strung together. It is to a denoising autoencoder what a deep-belief network is to a restricted Boltzmann machine. A key function of SDAs, and deep learning more generally, is unsupervised pre-training, layer by layer, as input is fed through.

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

An LSTM Autoencoder is an implementation of an autoencoder for sequence data using an Encoder-Decoder LSTM architecture. This example trains an autoencoder to generate text. The encoder uses a word embedding and an LSTM operation to map the input text into latent vectors.

1. **Explain Extractive summarization.**

Extractive summarization aims at identifying the salient information that is then extracted and grouped together to form a concise summary. Abstractive summary generation rewrites the entire document by building internal semantic representation, and then a summary is created using natural language processing.

1. **Explain Abstractive summarization.**

Abstractive summarization, on the other hand is a technique in which the summary is generated by generating novel sentences by either rephrasing or using the new words, instead of simply extracting the important sentences.

1. **Explain Beam search**

A beam search is most often used to maintain tractability in large systems with insufficient amount of memory to store the entire search tree. For example, it has been used in many machine translation systems. (The state of the art now primarily uses neural machine translation based methods.)

1. **Explain Length normalization.**

One simple length normalization formula is to divide the number of occurrences by the length of the document. For example, we can measure the length in pages and divide the number of occurrences (term frequency) by the number of pages.

1. **Explain Coverage normalization**

The plots are constructed by first determining the coverage of each individual base and normalizing the coverage over all the bases. To calculate normalized coverage, the coverage is divided by the average coverage over all 10 base positions.

1. **Explain ROUGE metric evaluation**

ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It is essentially a set of metrics for evaluating automatic summarization of texts as well as machine translations. It works by comparing an automatically produced summary or translation against a set of reference summaries (typically human-produced).