**APPROACH TO FIND THE SIMILARITY BETWEEN THE TEXTS**:

**Using deep learning technique that is**:

1. General model Architecture

2. A distance measure model

For Sentences, the model uses **pre-trained word embedding’s** to identify semantic similarities. Set of pre-trained word embedding’s utilized in this project. We had used is word2vec simple embedding’s form and to obtain those embedding’s used google News-Vectors .gz file .

which we can download from the below mentioned link.. <https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/edit>

**1**) Model architecture that we used is **Siamese LSTM** model consists of two identical subnetworks that compute some kind of representation vectors for two inputs

**Why?**

Reason to use Siamese LSTM is that it performs well on similarity tasks on texts as well as on images compared to other machine learning models. Word embedding’s trained with existing methods are not optimized for the sentence representation, whereas Siamese handles this problem by directly training and then averaging word embedding’s. This network learns word embedding’s by predicting the surrounding sentences from a sentence representation.

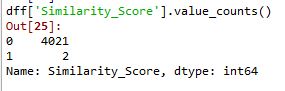
**2**) The **Manhattan LSTM** is which uses two LSTMs to measure similarity between a pair of sequences.

**Why?**

Manhattan distance slightly outperforms with other reasonable alternatives such as cosine similarity… Due to the curse of dimensionality, other distance methods become a poor choice as the number of dimensions increases.

***The model is trained on the questions dataset initially and those weights and word embedding’s of google news vectors were used to predict for any text sentences.***

***The model weights saved as SiameseLSTM.h5 and the similarity scores obtained for the Text\_Similarity\_Dataset.csv are stored in similarityscores.csv***

*** In this you can observe, found between two texts got as 0 represents not similar and 1 represents similar..4021 rows are not similar and only 2 rows are similar..***