

# **Introduction to NLP**

## **Assignment - 1**

### **Language Modelling**

Rishabh Khanna  
2019113025

## N-gram model:

An n-gram model says that the probability of a sentence is based on its count in the training corpus. Since we do not have unlimited data to get the exact possibilities, smoothing techniques prevent us from getting zero probability. We generally get zero probability when the test set has words that haven't been seen in the training set. To prevent these zero probabilities we can use mathematical smoothing techniques. IT relies on back-off methods that remove certain words from the history to get sequences we have seen in the train set.

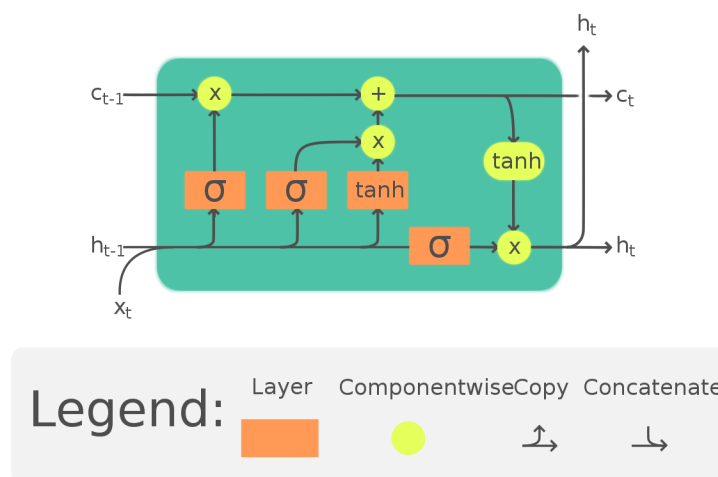
Witten Bell:

$$p_{WB}(w_i|w_{i-n+1}^{i-1}) = \lambda_{w_{i-n+1}^{i-1}} p_{ML}(w_i|w_{i-n+1}^{i-1}) + (1 - \lambda_{w_{i-n+1}^{i-1}}) p_{WB}(w_i|w_{i-n+2}^{i-1})$$

Kneser-Ney:

$$\hat{p}(w_i|w_{i-n+1}^{i-1}) = \frac{c(w_{i-n+1}^i) - D(c(w_{i-n+1}^i))}{c(w_{i-n+1}^i)} + \gamma(w_{i-n+1}^{i-1}) \hat{p}(w_i|w_{i-n+2}^{i-1})$$

## Neural Language model



Neural language models like LSTM is a better model as it has a lot more trainable parameters. LSTM is used for variable length sequences and can be given the history of

any length. We can use this to generate a probability distribution over the vocabulary and get probability and perplexity scores.

## **N-gram vs Neural Model**

- The perplexity score of neural models would be much lower than that of an n-gram model.
- Since we have a lot more trainable parameters the neural model would retain a lot more information about the corpus.
- Generally, a neural model can be used for much longer lengths and even variable-length seq.
- A neural model is better for prediction, as it would be much faster.

## **Perplexity Analysis**

- In the perplexities, I calculated the train perplexity is generally lower than the test perplexity this is because the model is over-tuned on the train data.
  - The model can be generalised by increasing the batch size while training.