Lab and Theory Assignment NLP

1. Represent the 2 equation of loss functions MSE and Cross Entropy in Computational Graph?
2. Gradient with respect to all the parameters of LSTM and GRU.
3. When will the LSTM and GRU will be similar to Vanilla RNN?
4. Compute back propagations step in self attention model?
5.

6.

7.

Assignment:

Vocab = $\{A, B, C, D, E\}$ $q_i \in \{t_1, t_2, t_3\}$

Tags : $\{t_1, t_2, t_3\}$

S1: AABBC   S2: BCDEA   S3: AABCD

S4: EEABC   S5: AEBCD

S1: $t_1 t_1 t_2 t_2 t_1$ S2: $t_2 t_3 t_1 t_2 t_1$ S3: $t_1 t_1 t_2 t_3 t_1$

S4: $t_2 t_1 t_1 t_2 t_3$ S5: $t_1 t_1 t_2 t_3 t_2$

(i) Compute λ : A, B, π (initial prob)

(ii) Assume 1st order HMM where t_i are hidden

(iii) Compute the probability of AAAABC | λ

(iv) Generate the most probable tag set for the following sequence of BBAAE.

Lab Assignment NLP

1. Generate Random data f(x) (0,1) for Y = 3*f(x) + 0.1 and train Linear regression model and plot Loss vs nitration plot, and learning rate and loss(each iteration)
2. Plot the function $(2-f(x))^2$, where f(x) = $(1-\text{sigmod function})^2$.
3. Crawl any corpus wiki around 10 million sentences and find average length of sentence, find the no of unique word in the crawled corpus.
4. Download any text corpus and represent your data using the following method:
   - BOW(Bag of word)
   - TF/TF-IDF
   - One hot vector representation
   - CBOW
   - SkipGram
5. Information extraction in the form of triplet of head, relation and tail of sheared Hindi dataset
6. Perform automatic extraction on the same dataset and use extracted triplets as training and testing data