Lec 4

- Bayes Rule:- $P(X|Y) = P(X)*P(Y|X) / P(Y)$
- The **chain rule** extends the Bayes rule to longer sequences.
- **Language model** ➔ is the statistical model of a language (probabilities of words).
- Language model can do word and letter prediction, can score and sort sentences, and also can do POS tagging.
- With the chain rule, the longer the sequence, the less likely we are to find it in a training corpus.
- With **Markov assumption,** the probability of the next word depends only on the previous K words.
- **N-gram** is the simplest model that assigns probabilities to sentences.
- N-gram probabilities come from a training corpus.
- The larger the 'n' in n-gram, the more the number of parameters to estimate.
- Unigram ➔ the probability of the sequence equals to the product of the probabilities of individual words.
  $P(w_1, w_2, ..., w_n) = P(w_1) * P(w_2) * ..... * P(w_n)$
- Bigram ➔ the probability of the next word depends only on the previous word.
  $P(w_1, w_2, ..., w_n) = P(w_1) * P(w_2| w_1) * ..... * P(w_n| w_{n-1})$
- Trigram ➔ the probability of the next word depends on the previous two words.
  $P(w_1, w_2, ..., w_n) = P(w_1) * P(w_2|w_1) *P(w_3|w_1,w2)...* P(w_n| w_{n-2},w_{n-1})$

- For n-gram model, we have ' n-1 ' start tokens <s> and only one end token </s>

- **Hidden Markov Model** is a special case of Bayesian inference, and has two assumptions:-
  1. The probability of a word appearing is dependent only on its own POS tag.
  2. The probability of a tag appearing is dependent only on the previous tag (bigram).

Lec 5

- **LOG probability** is used to solve the problem of underflow.
  Log(a*b) = log(a) * log(b)

- There are 2 ways of evaluating LMs:-
  1. Extrinsic ➔ in terms of external measure, and depends on some external task.
  2. Intrinsic ➔ in terms of properties of LM itself, and is independent on any task.
- With extrinsic evaluation, the utility of a language model is often determined in **practice.**
- **Perplexity (PP)** ➔ is the most common evaluation metric for n-gram LM, and it is the inverse probability of the test set, normalized by the no.of words.
- Minimizing the PP is equivalent to maximizing the test set probability.
- Lower PP means better model.
- Smoothing ➔ assign some non-zero probability to any n-gram, even one that was never observed in training.
- Smoothing make the distribution more uniform.
- Smoothing algorithms provide a better way of estimating the probability of n-grams.
- **Interpolation** ➔ mix the probability estimates from all the n-gram estimators.
- **Backoff** ➔ only to a lower order n-gram.
- Interpolation performs better on small training sets, backoff performs better on large training sets.

- **Morphology** ➔ is the study of the way words are built up from smaller meaningful units called: **morpheme**.
- **Morpheme** is the minimal meaningful unit in a language.
- **Morpheme parsing** ➔ breaking a word down into morphemes.
- Morphemes are 2 types:-
  1. Stems.
  2. Affixes.

- **Stem** ➔ is the core meaningful unit of a word.
- **Lemma** ➔ is an abstract conceptual form of a ward.
- **Stemming** ➔ is stripping off such word endings.
- **Affixes** ➔ add some additional meaning and grammatical functions to the words.
- Types of affixes:-
    1. Prefixes: precede the stem.
    2. Suffixes: follow the stem.
    3. Infixes: inside the stem.
    4. Circumfixes: precede and follow the stem.
- There are many ways to combine morphemes to create words:-
    1. Inflectional morphology: result in a word of the same class as the original stem, and doesn't change POS or the meaning of the word,
    2. Derivational morphology: result in a word of a different class and different meaning.
    3. Compounding: is the combination of multiple word stems together.
    4. Cliticization: is the combination of a word stem with a clitic.

- Building a morphological analyzer:-
    1. Lexicon: the first requirement and includes the list of stems and affixes.
    2. Morphotactics: explains the allowable morpheme sequences.
    3. Orthographic rules: used to model the changes occurring in a word when combining morphemes.
- **Morphological generation** is the opposite process of morphological analysis.