**Lecture 8 – Tagging**

**Goals**

1. Generative Tagging
   1. Parameterization
   2. Estimation
   3. Inference
2. Discriminative tagging
   1. Max Entropy
   2. Simple RNN
   3. Bidirectional RNN

**Generative Tagging**

* Sentence: I love white dogs
* Tags: Pr, Vb Adj, N
* Baseline: most frequent tag – already gets 90%+ accuracy
* We are given
* Parameterization
* Markov Assumption
  + - called the emission probability
    - called the transition probability
  + Number of parameters:
* Estimation
  + MLE of transition probability:
  + MLE of emission probability:
  + Remember there are smoothing issues
  + Want to use EM…but
* Inference Problem
  + We are looking for
    - Assume and are given
  + Solution: Viterbi Algorithm
    - Let be the length of the sequence
    - Goal:
    - Base Case: ,
    - Recursive Case
    - Complexity:
* Now we can do estimation with EM
  + Take a random guess of the parameters, and compute the MLE efficiently with the Viterbi algorithm
  + Bad results with random initialization
  + Can get good results with good initialization

**Discriminative Tagging**

* Now we estimate
* Max Entropy
  + - Indicator function of features
    - i.e.
* NN Structure for Max Entropy
  + Final layer: softmax, each unit represents
  + Vector representing indicator function for