**Lecture 4 – Expectation Maximization (EM)**

1. Topic Models
2. Parameter estimation for TM using EM
3. EM for MT (IBM Model 1)

**Clarification about Generative Models**

* Remember we have (different for each document) and (same for all documents)
* Assume we have topics and words
* Option 1
  + controls the proportion of topics
  + For ,
* Option 2
  + For ,

**EM**

* Extremely useful optimization procedure in NLP
* Consider the observed case
  + The observations are pairs
  + We can estimate   and
* What do we do in the unobserved case?
  + We want to maximize
  + We can use stochastic gradient descent or EM; today we discuss EM
  + For a given , we don’t know the – is unobserved.
    - Want to estimate
    - This is the expectation step – given my parameters, what is the distribution of the hidden variable?
    - Now for the maximization step
    - ,

**Application to MT**

* “the blue house” 🡪 “la maran blue”
* Alignment (records where each French word was in the English phrase)
* Observed Case:
  + We assume every word is translated independently
  + i.e.
  + Estimate the hidden variable :
  + Maximization:
    - Assume is uniform