HW5

* 1. With = 0, the second model is equivalent to the first one since the probability of choosing the background is 0 and we can ignore the background term.
  2. is the maximum likelihood estimate for this LM.
  3. Hypothesis: A large will make the topics indistinguishable from each other, which means that every topic has similar word probabilities. A small will make every topic have high probabilities for common words.

To test the hypothesis, we need to use a fixed collection of documents, a fixed background topic model and a fixed set of initial values as the starting point of EM algorithm. Then we can generate several large, small and normal values of and use them to run EM algorithm with other parameters stay the same. In the end, we can compare the results of topics generated by different values of .

1. Similar to the derivation in Chase’s note section 4.

E step:

Here dj denotes the jth word token in document d.

M step: