

Modeling a Multi-Topic Document

A document with 2 subtopics, thus 2 types of vocabulary

...

text mining passage

food nutrition passage

text mining passage

text mining passage

food nutrition passage

...

How do we model such a document?

How do we “generate” such a document?

How do we estimate our model?

We’ve already seen one solution – unigram mixture model + EM

This lecture is about another (better) solution – HMM + EM