

Summary

- PLSA = mixture model with k unigram LMs (k topics)
- Adding a pre-determined background LM helps discover discriminative topics
- ML estimate “discovers” topical knowledge from text data
 - k word distributions (k topics)
 - proportion of each topic in each document
- The output can enable many applications!
 - Clustering of terms and docs (treat each topic as a cluster)
 - Further associate topics with different contexts (e.g., time periods, locations, authors, sources, etc.)