

# Summary

- General behavior of a mixture model:
  - Every component model attempts to assign high probabilities to highly frequent words in the data (to “collaboratively maximize likelihood”)
  - Different component models tend to “bet” high probabilities on different words (to avoid “competition” or “waste of probability”)
  - The probability of choosing each component “regulates” the collaboration/competition between the component models
- Fixing one component to a background word distribution (i.e., background language model):
  - Helps “get rid of background words” in other component
  - Is an example of imposing a prior on the model parameters (prior = one model must be exactly the same as the background LM)