

The formal task

Given:

- Collection of texts as bags-of-words:
 n_{wd} is a count of the word w in the document d

Find:

- Probabilities of word in topics:

$$\phi_{wt} = p(w|t) \quad \longleftarrow \quad \text{Definition of a topic!}$$

- Probabilities of topics in documents:

$$\theta_{td} = p(t|d)$$