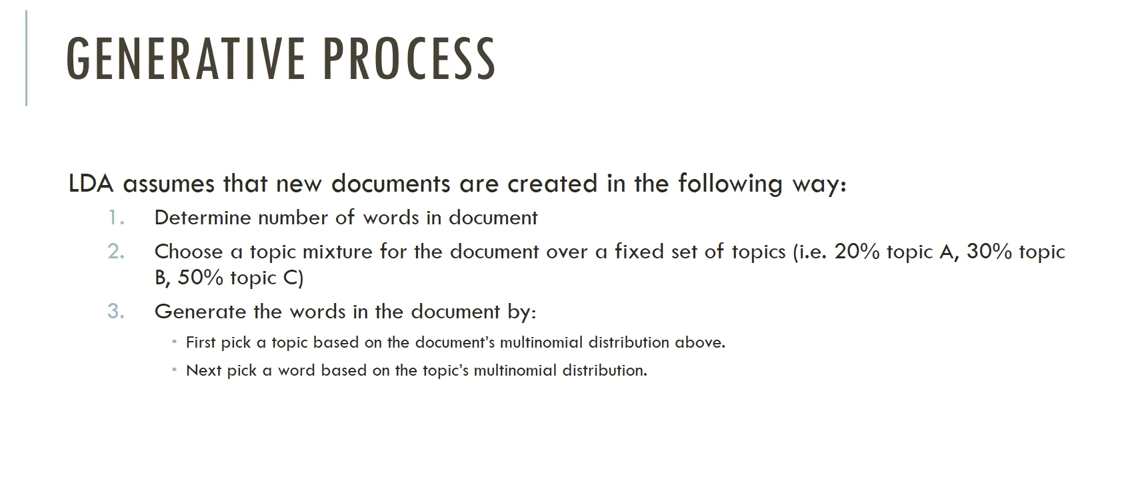
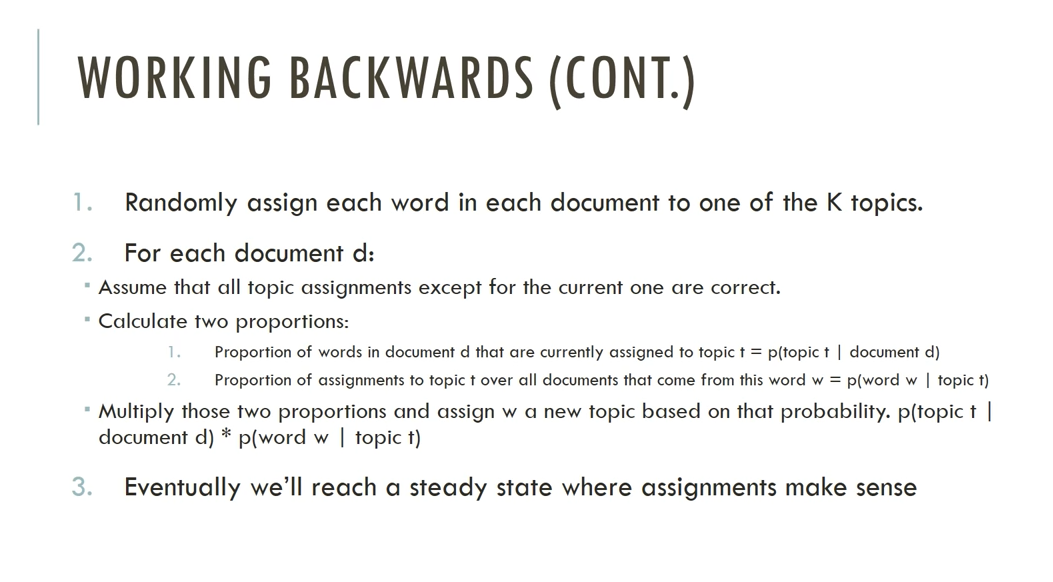
LDA Emoji Demo and Associated Article:

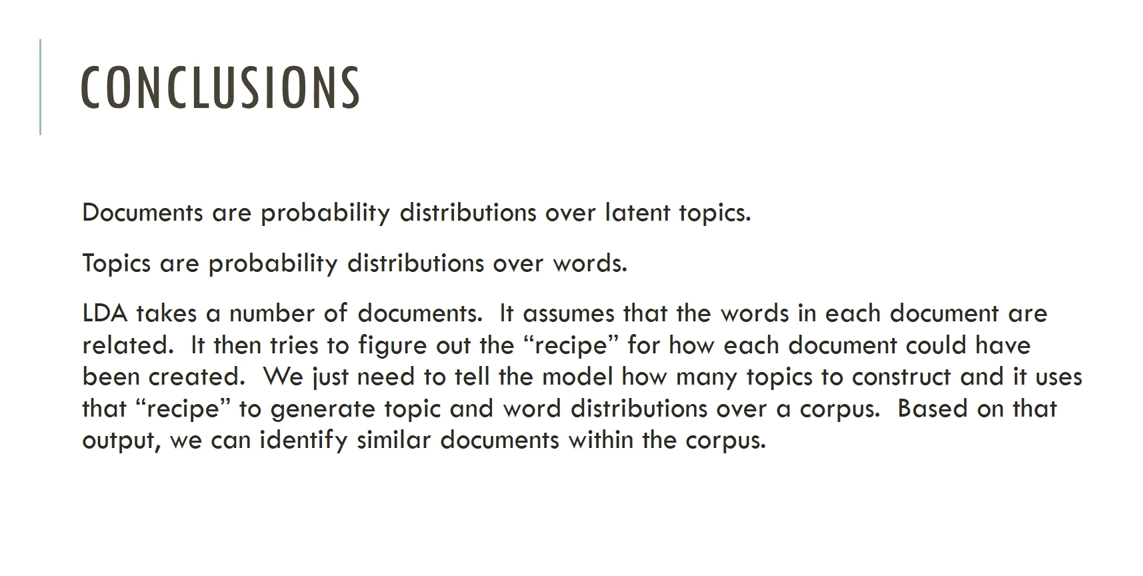
<https://medium.com/@lettier/how-does-lda-work-ill-explain-using-emoji-108abf40fa7d>

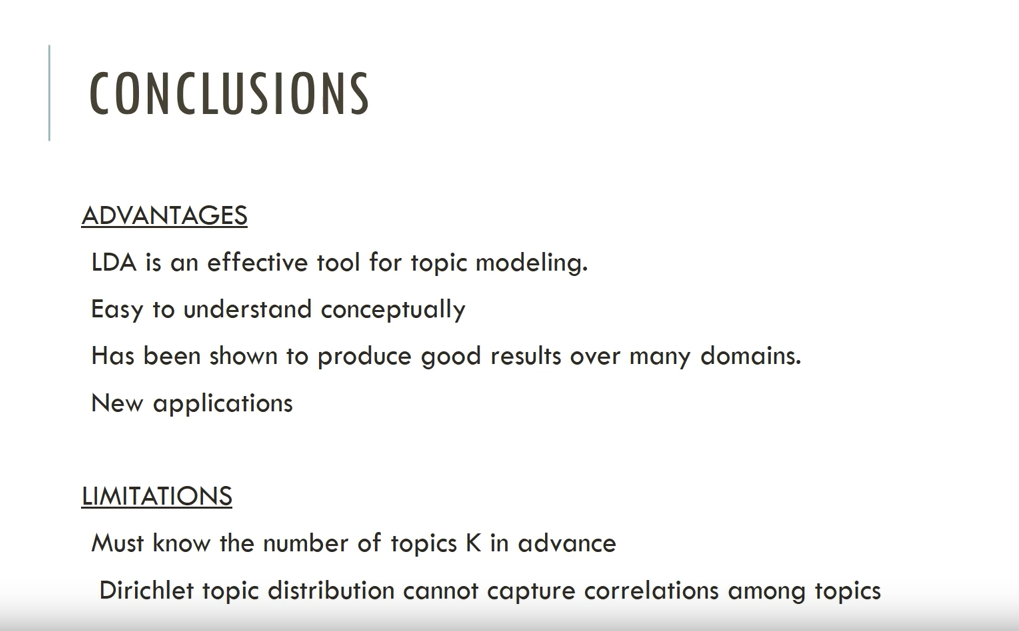
How does LDA generative process work?











Latent Dirichlet Allocation (LDA) Hyper-parameters

Alpha and Beta represent the Dirichlet Priors

Alpha: sets the prior on the per-document topic distributions

- High alpha means every document is likely to contain a mixture of most of the topics and not primarily fall into a single topic

- Low alpha means every document is likely to be represented by just a few of the topics

Beta: sets the prior on the per-topic word distributions

- High beta each topic likely to contain a mixture of most of the words and not primarily just a few of the words

- Low beta means each topic is likely to be represented by just a few of the words in the corpus

High alpha will make documents appear more similar to each other

High beta will make topics appear more similar to each other

LDA model at inference time ignores any word that is not used in the training set.

Let the right words in:

- remove very short or possible very long words

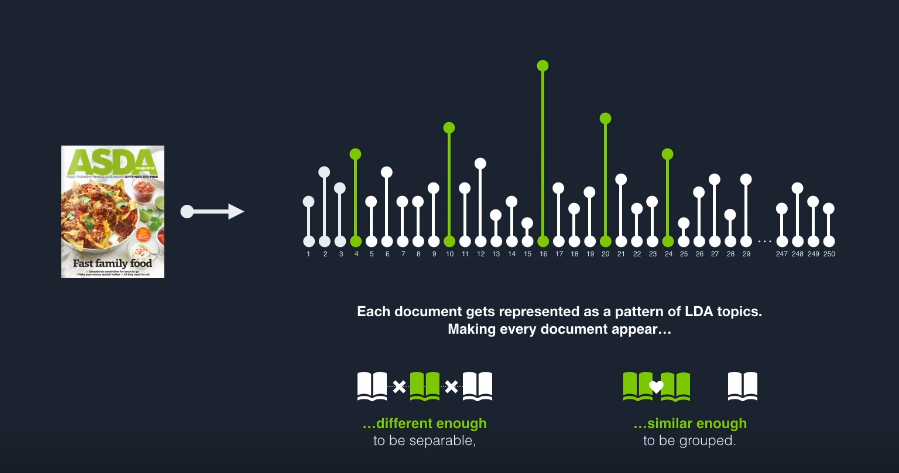
- remove very rare and very common words

- remove stop words

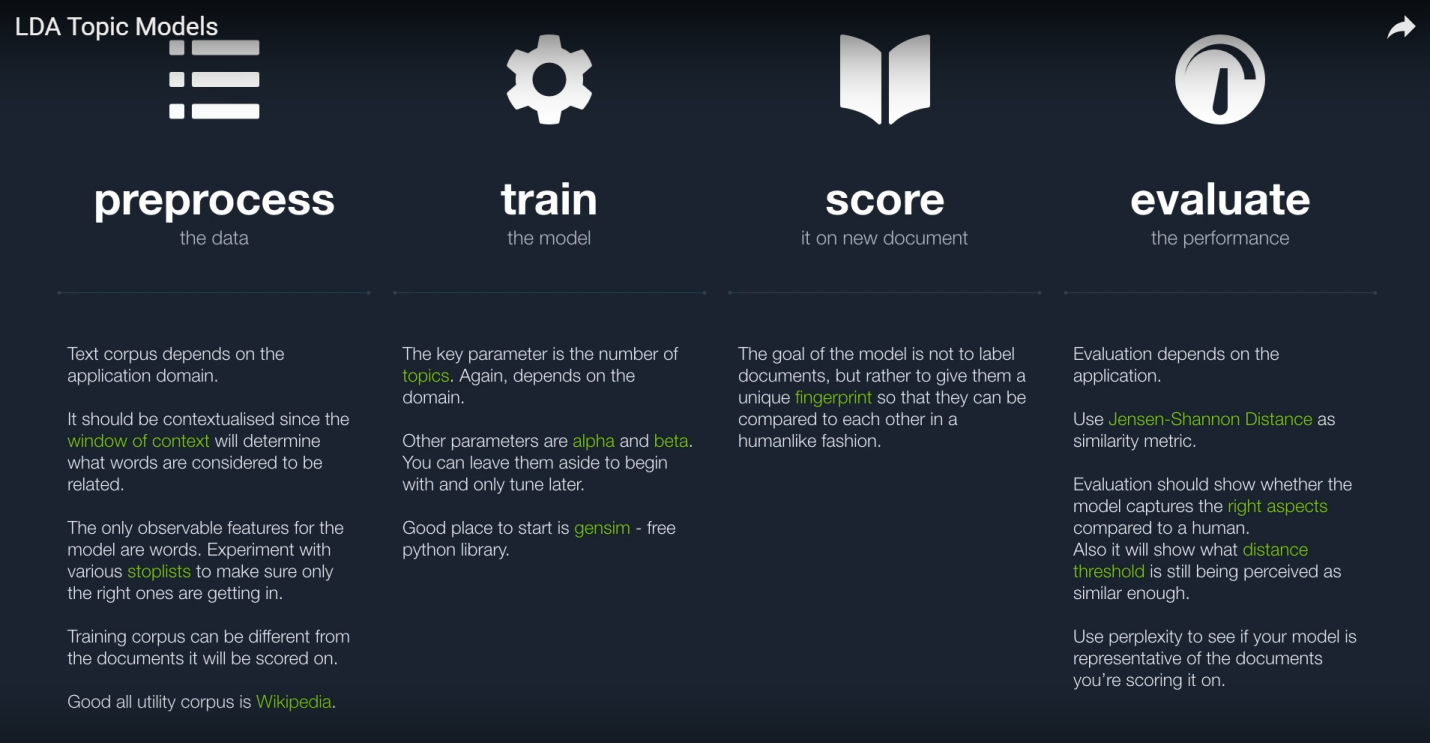
- consider lemmatization

- part of speech tagging (nouns/verbs tend to be more informative)

- keep top n words (recommended between 50k and 100k words for a large text corpus - Wikipedia scale)



Jensen-Shannon Distance for measuring the pairwise similarity between two probability distributions.



"All models are wrong, but some are useful" - George E. P. Box

Good reference video: <https://www.youtube.com/watch?v=3mHy4OSyRf0>