Department of Computer Science
The University of Melbourne
COMP90042 WEB SEARCH AND TEXT ANALYSIS (Semester 1, 2018)

Workshop exercises: Week 9

Discussion

1. In modelling terms, what is the difference between **topics** and **classes**?
2. What is a **topic model**? What is the difference between topic modelling and text classification?
3. Give 3 example applications for topic models. Explain why it is not feasible to use text classification for these applications.
4. It is possible to train a topic model using unsupervised HMMs but this is not ideal. Why? How it can be improved?
5. How can you evaluate topic models automatically?
6. Cite 2 example visualisations for evaluating topic models manually.
7. Cite 3 extensions of LDA and what kind of problems they address.

Programming

1. Go through the `WSTA_N15_topic_models` notebook. What kind of topics do you get in your final output? Can you label all of them? How would you improve the interpretability of these topics?

Catch-up

- What is a language model?
- What is the difference between n-gram LMs and neural LMs?
- How do you evaluate language models?

Get ahead

- Try some of the extensions proposed in the `WSTA_N15_topic_models` notebook.
- Try the Gensim tutorial on finding topics on Wikipedia (`https://radimrehurek.com/gensim/wiki.html`). Beware though: training on Wikipedia can take quite a long time.