School of Computing and Information Systems
The University of Melbourne
COMP90042 NATURAL LANGUAGE PROCESSING (Semester 1, 2020)

Workshop exercises: Week 11

Discussion

1. What is **Question Answering**?
   (a) What is **semantic parsing**, and why might it be desirable for QA? Why might approaches like NER be more desirable?
   (b) What are the main steps for answering a question for a QA system?
2. What is a **Topic Model**?
   (a) What is the **Latent Dirichlet Allocation**, and what are its strengths?
   (b) What are the different approaches to evaluating a topic model?

Programming

1. In the iPython notebook `12-topic-model`, we build a topic model on the Reuters news corpus.
   - Explore different number of topics: qualitatively how does it change the topics?
   - Explore different values of the document-topic α and topic-word η (β in lecture) priors: qualitatively how does it change the topics? What values work best for the downstream document classification task? (Note: you can also try 'auto' where the model will try to learn these hyper-parameters automatically)
   - Modify the classification task such that it uses bag-of-word *and* the topic distribution as input features to the classifiers. Do you see a performance gain?