

Data Science Toolbox Portfolio Questions

08 Topic Model and LDA

Daniel Lawson — University of Bristol

Block 8

Portfolio 08

Choose **one question** and write up to **one page** about it. You are free to conduct further experiments to add weight to your results, and any additional material you generate can be submitted as an appendix. See [The Assessment Page](#) for advice.

These questions may make reference to the content from the current block.

Question R08.1: Consider the paper [Exploring Topic Coherence over many models and many topics](#). Explain and motivate the models for coherence, and critique the conclusion that “LDA best learns descriptive topics while LSA is best at creating a compact semantic representation of documents and words in a corpus”.

Question R08.2: In the workshop, LDAvis presented both “relevance” [LDAvis: A method for visualizing and interpreting topics](#) and “saliency” [Termite: Visualization Techniques for Assessing Textual Topic Models](#). Examine the documentation for these terms and explain when one could be preferred over the other.

Question R08.3: From [Eric Jang’s “A Beginner’s Guide to Variational Methods”](#) or otherwise, explain what the KL Divergence between distributions is and how it relates to Variational Inference. What barriers are there to doing Variational inference in practice?