

# STCS 6701: Foundations of Graphical Models: Reading 8

Maxime TCHIBOZO (MT3390)

November 2020

## 1 Supervised Topic Models (2008) - McAuliffe and Blei

LDA can be a useful tool for dimensionality reduction and feature extraction when the topic proportions are used for a downstream classification task. However, while the identified unsupervised topics capture statistical properties of the corpus, they are unlikely to be meaningful for (or well-suited to) the given classification task.

Supervised Topic Models (sLDA) pair each document with a response, and aim to infer latent topics predictive of the response. Thus, sLDA favors topics which will lead to satisfactory performance on the supervision task (movie review ratings, web page popularity), with the expectation that the topics will learn to group similarly meaningful (relatively to the supervised task) words together.

In sLDA, documents  $x_i$  and their response  $y_i$  are modelled jointly, and this joint distribution can be derived by using a Data Generating Process which resembles that of LDA:

1. Draw topic proportions  $\theta \mid \alpha \sim Dir(\alpha)$
2. For each word:
  - (a) Draw topic assignment  $z_n \mid \theta \sim Mult(\theta)$
  - (b) Draw word  $w_n \mid z_n, \beta_{1:K} \sim Mult(\beta_{z_n})$
3. Draw response variable  $y \mid z_{1:N}, \eta, \sigma^2 \sim \mathcal{N}(\eta^T \bar{z}, \sigma^2)$   
with  $\bar{z} = \frac{1}{N} \sum_{n=1}^N z_n$

The response is non-exchangeable in this case, and appears downstream of the sentence generation process.

This paper details the mathematical derivations for Variational Inference (VI) in the special case of an  $\mathbb{R}$ -valued response  $y$ , then generalizes VI to any response generated by a Generalized Linear Model (GLM).

From a practical viewpoint: to assess convergence, the authors run VI until the change in the corpus-level likelihood bound is less than 0.01%, and the relative change in the per-document ELBO is less than 0.01%. They use  $R^2$  with 5-fold cross-validation to compare sLDA to Linear Regression (LR) with unsupervised LDA topics.

They find sLDA to be superior to both LDA + LR and LDA + Lasso. It would have been interesting to visualize the contents of topics, to confirm or reject our intuition that sLDA topics should be more meaningful than unsupervised LDA topics.