Posterio Collapse and Latent Variable Non-identifiability

Yixing Wang
University of Michigan.

3 February 2022, 2.30pm UK time

Abstract
Variational autoencoders model high-dimensional data by positing low-dimensional latent variables that are mapped through a flexible distribution parametrized by a neural network. Unfortunately, variational autoencoders often suffer from posterior collapse: the posterior of the latent variables is equal to its prior, rendering the variational autoencoder useless as a means to produce meaningful representations. Existing approaches to posterior collapse often attribute it to the use of neural networks or optimization issues due to variational approximation. In this paper, we show that posterior collapse is a problem of latent variable non-identifiability. We prove that the posterior collapses if and only if the latent variables are non-identifiable in the generative model. This fact implies that posterior collapse is not a phenomenon specific to the use of flexible distributions or approximate inference. Rather, it can occur in classical probabilistic models even with exact inference, which we also demonstrate. Based on these results, we propose a class of identifiable variational autoencoders, deep generative models which enforce identifiability without sacrificing flexibility. This model class resolves the problem of latent variable non-identifiability by leveraging bijective Brenier maps and parameterizing them with input convex neural networks, without special variational inference objectives or optimization tricks. Across synthetic and real datasets, identifiable variational autoencoders outperform existing methods in mitigating posterior collapse and providing meaningful representations of the data.

References

[1] Y. Wang, D. Blei, J.P. Cunningham, Posterior Collapse and Latent Variable Non-identifiability, NeurIPS 2021.