

CS 294 Project Proposal:

Variational Auto-encoders with Applications to Sequential Data

Project members: Jianbo Chen Cheng Ju Billy Fang

26 September 2016

Team members

- Jianbo Chen (Statistics PhD student)
- Cheng Ju (Biostatistics PhD student)
- Billy Fang (Statistics PhD student)

Problem statement and background

We would like to focus on generative models for tasks in the sequential data setting, with an emphasis on variational auto-encoders (VAEs). First, we plan to start from implementing existing models which incorporates VAE into the sequential data. For example, VAE ideas are incorporated into recurrent neural network (RNN) applications to generate sequential data like speech and handwriting. (Gregor et al. and Chung et al. 2015) Then we would like to investigate whether we can improve the current methods, for example, by borrowing ideas from another popular generative model: generative adversarial nets (GANs). We may also design new variants/generalizations of VAEs in the hope of overcoming the drawbacks of the popular VAE model (detailed below).

Replication

First, we would like to implement several variational auto encoders using TensorFlow in order to familiarize ourselves with the framework and the workflow. Concretely, we would like to implement the Auto-encoding Variational Bayes model as proposed in [5], which trains it on the images from the MNIST and Frey Face datasets. We also will combine VAEs with RNNs on sequential data that generates speech and handwriting, proposed by [4] and [1] with the hope of improving the results by modifying the neural network structure in the model.

Drawbacks of VAE and potential solutions

Below we list several of the potential problems of VAE and also draft our untested solutions. We hope to make these conjectured models more concrete, implement them on MNIST and Frey Face datasets, and see their effects when carrying out the project.

- Images generated from variational auto-encoders trained on images tend to be somewhat blurry. [3] On the other hand, GANs tend to produce sharp images but are difficult to train. We would like to investigate whether it is possible to incorporate adversarial ideas from GANs to sharpen the output of VAE generative models.

- Another troubling issue with contemporary VAE models is that they tend to use only a small subset of the dimensions of the latent variable z [3]. This indicates that increasing the dimension of z will not actually increase the effective expressiveness of the model. Rather than increasing the dimension of z , we consider more complicated graphical models. In particular, there are many hierarchical models that contain more complicated probabilistic structure and expressiveness than the vanilla VAE. Can we gain from adopting the expressive power of more complicated hierarchical models, while preserving the simplicity of inference/computation in the variational auto-encoder?

Applications

Following the work of [4] and [1], we aim to combine VAEs with RNNs to generate sequential data such as handwriting or speech. In the latter paper, Chung et al. argued that through the use of high-level latent random variables, the variational RNN (VRNN) can model the kind of variability observed in highly structured sequential data. Beyond being able to generate new sequential data, the success of these models lies in their ability to learn the features and processes that generated the observed data. By studying and improving these models, we hope to gain a better understanding of the essential components of speech and handwriting generation, and perform tasks like recovering missing data or predicting/completing a sequence of data. If time permits, we may consider other image generation tasks in general.

Source of data

We will mainly use MNIST handwriting dataset and Frey Face datasets. We will also try to apply the model to CIFAR 10 and ImageNet.

For the RNN applications, we will consider sequential datasets, like the Blizzard Challenge 2013 text-to-speech dataset and the IAM-OnDB handwriting dataset.

Tools

We plan to implement the models in TensorFlow. For hardware, we plan to use GPUs from Amazon AWS. We may also refer to open source code, including the following.

- <https://github.com/ikostrikov/TensorFlow-VAE-GAN-DRAW>
- <https://github.com/ericjang/draw>

Evaluation

One natural approach is to evaluate a generative model using the log likelihood of test datasets. However, as argued in [7], a good log likelihood is neither necessary nor sufficient for producing plausible samples. We will thus consider other evaluation metrics like Parzen windows and visual evaluation of samples. Ultimately our choice of evaluation metric will depend on the particular application at hand.

References

- [1] Junyoung Chung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Aaron C Courville, and Yoshua Bengio. A recurrent latent variable model for sequential data. In *Advances in neural information processing systems*, pages 2980–2988, 2015.
- [2] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.
- [3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. Book in preparation for MIT Press, 2016.
- [4] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*, 2015.
- [5] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [6] Marcus Liwicki and Horst Bunke. Iam-ondb-an on-line english sentence database acquired from handwritten text on a whiteboard. In *Eighth International Conference on Document Analysis and Recognition (ICDAR’05)*, pages 956–961. IEEE.
- [7] Lucas Theis, Aäron van den Oord, and Matthias Bethge. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*, 2015.