Variational Auto Encoders

A deep learning approach to dimensionality reduction and generative modelling

Una Singo
Marcus Gawronsky

September 2019

Abstract

Variational Autoencoders (VAEs) are a class of generative artificial neural network commonly used for dimensionality reduction in machine learning applications. In this paper, we benchmark their application against traditional methods in multivariate statistics, with their aim of applying this method to a common open dataset.

Literature Review

Variational Autoencoders (VAEs) are a class of generative artificial neural network commonly used for dimensionality reduction in machine learning applications. The technique extends on the application of traditional autoencoders and deep belief networks by introducing techniques from Variational Bayesian Methods.

An autoencoder consists of two neural networks, an encoder network and decoder network. The encoder network takes an input and produces a lower-dimensional representation of the data. The decoder network takes the lower-dimensional representation of the data and learns a reconstruction of the data back into its original vector-space. These neural networks can have a varying number of hidden-layers, activation functions and neurons to learn non-linear representations of the data [Bengio et al., 2013, 2007].

Variational Bayesian Methods are a class of techniques used in approximating intractable integrals found in Bayesian Modeling. Using an approximate distribution, $q^{\star}(x)$, a loss function is used to minimize the difference between the approximate distribution and true posterior. This loss function is often chosen to be the Evidence Lower Bound (ELBO) or Kullback–Leibler divergence and is weighted in the optimization procedure to ensure the validity of the applied approximate distribution.
The Variational Autoencoder aims to extend on encoder-decoder model architectures by placing distributional assumptions on the low-dimensional latent space. By making this assumption, these models can be generative, allowing one to sample data across a complex data manifold. While several methods can be used to ensure distributional assumption, the original authors include the Kullback–Leibler divergence into the loss function of the model to ensure a distribution over latent variable which is approximately standard normal [Kingma and Welling, 2013; Tolstikhin et al., 2017; Zhao et al., 2017].

Using these techniques, Variational Autoencoders provide a flexible method which can be applied to solve a wide range of problems across domains. In the original paper, Kingma and Welling [2013] use the popular MNIST dataset to both embed and generate new images of handwritten digits. A significant number of subsequent research on VAEs has focused on the learning latent distributions of images and generating new images. Generative Adversarial Networks are a popular example of such research outputs where the objective is to generate fake images that appear to be authentic to human observers [Goodfellow et al., 2014].
Apart from image generation VAEs have been shown to solve anomaly detection problems. An and Cho [2015] propose a method using the reconstruction probability from a variational autoencoder to identify outliers. Their experimental results show that this method outperforms traditional auto-encoder and principal components based methods.

References

An, J. and Cho, S. (2015). SNU Data Mining Center 2015-2 Special Lecture on IE Variational Autoencoder based Anomaly Detection using Reconstruction Probability.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2007). Greedy Layer-Wise Training of Deep Networks Yoshua. *In Advances in neural information processing system*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Nets. page iii.

Kameoka, H., Kaneko, T., Tanaka, K., and Hojo, N. (2019). ACVAE-VC: Non-Parallel Voice Conversion with Auxiliary Classifier Variational Autoencoder. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 27(9):1432–1443.

Kingma, D. P. and Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv e-prints*, page arXiv:1312.6114.

Tolstikhin, I., Bousquet, O., Gelly, S., and Schoelkopf, B. (2017). Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*.

Zhao, S., Song, J., and Ermon, S. (2017). Infovae: Information maximizing variational autoencoders. *arXiv preprint arXiv:1706.02262*.

Appendix

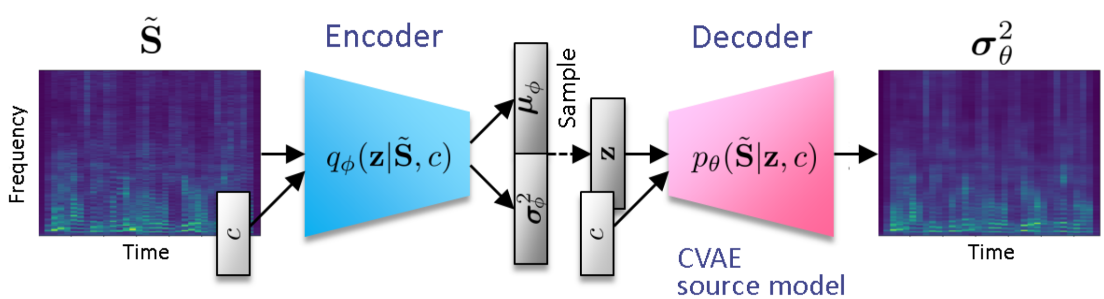

Figure 1: Diagram of how the data flows through the VAE Kameoka et al. [2019]