Literature Review - Computational Graphs for Machine Learning: Midterm Project
Autumn 2019
ECE471

General Overview & Purpose

This literature review serves to summarize papers concerning so-called beta variational autoencoders (β-VAE). The goal is to understand the generative capabilities of variational autoencoders(VAE), first introduced by Kingma and Welling [1] , and how β-VAE, a model published by DeepMind [2] [3], improves the ability of the network to yield disentangled representations.

Paper 1

Title: Auto-Encoding Variational Bayes [1]
Authors: Kingma, Welling - 2014
Notes and Discussion:

This is the paper that introduced the notion of a VAE. First, they show how to optimize a sampling process via the reparameterization trick. They then showed that intractable posterior inference can be made by fitting an approximate inference model using the proposed lower bound estimator. This varational lower bound forms the objective function.

Effectively, each input datapoint x produces a distribution over the possible values of a latent code z, viewed as an *encoder*, and each code word z produces a distribution over the possible corresponding values of x, forming a *decoder*.

Paper 2

Title: β-VAE: Learning Basic Visual Concepts with a Constrained Variaitonal Framework [2]
Authors: Higgins, et.al- 2017
Notes and Discussion:

This was the first paper released by DeepMind concerning β-VAEs. Their main contribution was to add a hyperparameter β to the objective function discussed in [1]. This additional parameter allows for explicit tuning of the ability of the network to learn a disentangled representation. Effectively, increasing β increases the latent channel capacity at the cost of reconstruction accuracy. β-VAE outperforms other generative models, both unsupervised (InfoGAN) and semi-supervised (DC-IGN), in terms of it's ability to learn disentangled factors.

The authors don't go into great detail about the network architecture itself, nor do they explicitly justify why β-VAE outperforms other SOTA models beyond qualitative remarks about 'disentanglement.' However, they go into far more detail in the companion paper released a year later, [3]

Paper 3

Title: Understanding disentangling in β-VAE [3]
Authors: Burgess, et.al- 2018
Notes and Discussion:

This paper, released by DeepMind a year later, takes the time to explain how and why β-VAE sets the SOTA for generative learning of disentangled representations. The authors define what they mean by a disentangled representation, and walk through the derivation of the objective function discussed in [1]. They then explain why this hyper-parameter β improves the network in terms of mutual information.

Effectively, the parameter β finds latent components which make different contributions to the log-likelihood term of the objective function in [1], and that latent components correspond to features that are qualitatively different. Furthermore, forcing a diagonal correlation matrix as part of the reparameterization trick also encourages the latent dimensions to align with generative factors. They then add an additional factor to the loss function to improve the effective 'channel capacity', as it were, of the network, which in turn increases the models ability to generate disentangled features. They also present the network architecture in the appendix for both the encoder and decoder networks.

References

[1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *CoRR*, vol. abs/1312.6114, 2013.

[2] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," in *ICLR*, 2017.

[3] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in beta-vae," *ArXiv*, vol. abs/1804.03599, 2018.