Development of the Gamma-VAE with applications in computer vision

Master thesis summary

Péter Tamás Kovács

June 2, 2023

In this contribution, we investigate the Gamma-VAE, a variational autoencoder (VAE) that uses Gamma distributions as prior and posterior. Albeit hardly explored, we maintain that this model – due to its nonnegative activations in the latent space – could model the neuronal system better than previous models not imposing this constraint. On the other hand, it could also allow for the investigation of the effects of nonnegativity on the emerging latent representations.

Our main contributions are twofold: first we elaborate on training stability regarding the model. We observe that its training is plagued by numerical instabilities that may be alleviated by proper parameterization of the posterior distributions and appropriate activation functions. Even though our findings may not be applicable to every architecture and dataset, we believe our considerations provide a useful starting point for later investigations.

Secondly, we analyze the properties of two Gamma-VAE architectures employed on different datasets (3DShapes, the Van Hateren dataset). By re-creating experiments already published with different VAEs, we aim to compare our Gamma-VAE with earlier, more established models at properties that are sought after in computational neuroscience: disentanglement and sparsity. In other words, our main focus are the properties of the latent distributions that are formed in the models. With the 3DShapes dataset, we probe the disenanglement of the representations by the well-known MIG score. We find that our model shows competitive results. On the Van Hateren dataset, we assess the sparsity of emerging representations by considering the emerging filters (vectors) in a linear decoder. We find that with proper regularization, our model exhibits 'Gabor-like' (localized, both spatially and in frequency) filters which are thought to be a good indicator of sparsity.

We thus conclude that our Gamma-VAE model has relevance for computational neuroscience, since it reaches competitive performance by applying the nonnegativity constraint of the neuronal system by design – a relatively novel feature. Obviously, to better understand the properties of the model and its potential applications, further research is needed. We believe that our contribution is a good foundation for that.