Notes: "Avoiding Latent Variable Collapse with Generative Skip Models"

February 2020

Summary and background:

1. VAEs model high-dimensional data by projecting inputs onto a lower dimension latent space
2. This is done in practice by training two separate models; $q_\phi(\mathbf{z_i}|\mathbf{x_i})$, and $p_\theta(\mathbf{x_i}|\mathbf{z_i})$, also known as the encoder and decoder respectively.
3. A problem can occur called "latent variable collapse" where the posterior, $q_\phi(\mathbf{z_i}|\mathbf{x_i})$, is approximately equal to the prior, $p(\mathbf{z})$; this causes the latent space representation to be very low in information.
4. This paper recommends skip connections from the latent space into the decoder to reinforce corrolation between latent variables and the recreation.

More details: An important task for VAEs is modeling $p_\theta(\mathbf{x}, \mathbf{z}) \equiv p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})$. Typically, the prior $p_\theta(\mathbf{z})$ is simply a unit guassian $\mathcal{N}(0, I)$, and the likelihood $p_\theta(\mathbf{x}|\mathbf{z})$ is a deep neural network parameterized by θ.

Consider the approximation $q_\phi(\mathbf{z_i}|\mathbf{x_i})$ modeled by a deep neural network parameterized by ϕ; the neural network takes in x_i and predicts z_i.

The goal of the VAE is to optimize the evidence lower bound (ELBO):

$$ELBO := \sum_{i=1}^{N} [E_{q_\phi(\mathbf{z_i}|\mathbf{x_i})}[\log p_\theta(\mathbf{x_i}|\mathbf{z_i})] - \mathbf{KL}(q_\phi(\mathbf{z_i}|\mathbf{x_i})||p(\mathbf{z_i}))] \tag{1}$$

The VAE optimizes (1) jointly; optimizing θ produces a good generative model, and optimizing ϕ produces a good approximate posterior.

Latent variable collapse happens when $q_\phi(\mathbf{z_i}|\mathbf{x_i}) \approx p(\mathbf{z_i})$; when the output of the posterior approximator q_ϕ does not depend on the data. This is a pathological state that does not necessarily produce perfect recreations, however, the state *is* a local minimum of sorts (or else it obviously wouldn't stay in it). In some cases, it may take the mean or median value, or in the case of images, may produce "blotches" that correspond to the lowest (or near) loss that can be achieved by "averaging" values.

Skip-VAE: The authors of Skip-VAE (henceforth SVAE) consider representing the ELBO instead as:

$$E_{p(\mathbf{x})}\{E_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})]\} - \mathcal{I}_q(\mathbf{x},\mathbf{z}) - \mathbf{KL}(q_\phi(\mathbf{z})||p(\mathbf{z})) \tag{2}$$

$$\mathcal{I}_q(\mathbf{x},\mathbf{z}) := E_{p(\mathbf{x})}E_{q_\phi(\mathbf{z}|\mathbf{x})}\log q_\phi(\mathbf{z}|\mathbf{x}) - E_{q_\phi(\mathbf{z})}\log q_\phi(\mathbf{z}) \tag{3}$$

(3) is the mutual information between $\mathbf{x}$ and $\mathbf{z}$.

The ELBO in (2) reveals that setting KL $= 0$ is equivalent to setting $\mathcal{I}_q(\mathbf{x},\mathbf{z}) = 0$

The authors note that our goal is not merely preventing the KL from equaling 0; a method of doing such is merely building an autoencoder.

The basis of SVAE is that if one adds residual paths from the latent space into the deeper layers of the decoder, $p_\theta(\mathbf{x}|\mathbf{z})$, you achieve better mutual information.

This is almost self-evident; by connecting more layers to the latent space, it will have a stronger effect. If the connections are weighted, as long as the weights to not all anneal to 0, the latent space will have more overall effect on the output, which is what we desire when building VAEs.

Does it work in real life? Eh, sorta. It's not groundbreaking. However, considering the ease of adding it to an existing VAE (couple lines of code), and the fact that posterior collapse *is* a common problem, it's a worthwhile technique to know about.