Project Writeup

Ryan Pyle

December 2018

1 VAE Latent Space

VAEs transform an input x into some latent space z. z is often given a independent Gaussian prior, allowing new samples to be generated by sampling from $N(0, I)$. How useful is this latent space? Does setting an independent Gaussian prior on our latent space cause issues?

The second issue is easier so I examine it first. It turns out that many latent variables may not map naturally to a $N(0, I)$. However, mapping to another distribution may not fix the issue - mapping to other distributions is harder, and we still need to know in advance what a 'good' target distribution is. Instead, several recent papers have examined the concept of normalizing flow, most notably the NICE, inverse auto-regressive flow, and GLOW frameworks. The key shared idea is that we can map to a latent $N(0, I)$ distribution, but then add an extra network section that acts as a paramaterized transformation of the latent probability mass. This learned transformation allows for the $N(0, I)$ to be flexibly made into a more realistic or useful latent space, improving outputs, while still keeping the essential quality of generation via $N(0, I)$ sampling.

The other issue is that latent spaces may be non-interpretable. We would prefer that latent spaces were easy to understand, e.g. a latent variable mapped to a single concept of interest. This can be promoted by having a disentangled latent representation, assuming that generating factors are independent. In VAE, this can be accomplished via β-VAE. β-VAE is a method that weights the second, KL divergence term of the standard VAE formula. This puts greater emphasis on having a $N(0, I)$ representation, which promotes disentanglement since each dimension is independent. The exact Gaussian shape may be more of an issue, but the NICE / GLOW frameworks could help solve this. In the β-VAE paper, it is markedly better at discovering and disentangling factors than either standard VAE or the top GAN disentangler, INFO-GAN.

Finally, latent space are more complex than they appear at first glance. In order to use them to their full efficiency, we need to take into account the underlying geometry or curvature of latent space. This essentially means that points with low data (that are e.g. unrealistic) are harder to travel through, acting as an increased distance in latent space. This has implications for e.g. interpolation where you need to interpolate through areas of low curvature rather than directly going from point A to B in latent space, if there is a high curvature region between them.

2 RNNs and Predicting in Latent Space

Another possibly utility of latent space is prediction. Given the lower dimensionality of a latent space, they should be easier to predict from. Recent works in RNN have shown that in many real-world scenarios they act as excellent function approximators for a variety of problems, making them an excellent candidate to use.

My project was to use a β-VAE on a simple system to get low dimensional, informative latents. I trained a RNN on the latent data generated, and used it to predict future latent states. The latent states were then fed back through the β-VAE decoder, giving final predicted outputs.

The original input was a bouncing ball of light in a small (28 by 28) grid, where the ball was a point that bounced without loss of momentum, and projected light in a small radius creating an image. The process was fully deterministic given initial conditions and velocity. My β-VAE had both decoder and encoder as a 2 layer fully connected network, with the latent dimensionality equal to 3 (the smallest that still gave accurate reconstructions). These 6 latent dimension (3 mean, 3 variance) were fit to a standard RNN until the reconstructions looked at least approximately accurate (e.g. similar scales and frequencies). Final reconstructions successfully recreated a ball of light that approximately represented the initial inputs. Strong points were that the output bounced off walls, and approximately conserved momentum. Downsides were that the ball occasionally morphed from point to point, and often varied its direction over time rather than continuing on in a straight line.