Notes: "Adversarial Autoencoders"

Sun

March 23, 2020

1 Notes:

1.1 Summary:

1. Instead of using the KL divergence (KLD) between a prior $p(\mathbf{z})$ and the variational posterior $q(\mathbf{z}|\mathbf{x})$, adversarially match the two using positive and negative sampling. This creates a model called an AAE.
2. Similar to VAEs, but the adversarial loss function allows arbitrary priors to be used.
3. $q_\phi(\mathbf{z}|\mathbf{x})$ is no longer constrained to specifying particular posteriors, and can even be a vanilla autoencoder encoder that is forced to be mapped to $p(\mathbf{z})$.

1.2 Medium level:

When it comes to autoencoders, one almost always wants a good continuous latent space. VAEs solve this problem by forcing the latent space roughly to approximate a given continuous prior, $p(\mathbf{z})$, usually a unit Gaussian. Traditionally, one just adds a regularization loss term to the existing autoencoder reconstruction loss term. This regularization term is usually the KLD between the variational posterior, $q(\mathbf{z}|\mathbf{x})$, and the prior, $p(\mathbf{z})$. For distributions without a closed for solution for the KLD and without computationally efficient estimators, this becomes intractable.

A solution for this is simple. Build a discriminator, $d_\psi(\mathbf{z})$, that takes positive samples (samples from the true distribution $p(\mathbf{z})$), and negative samples (samples from the aggregate posterior, $q(\mathbf{z}|\mathbf{x})$). By iteratively updating the model parameters θ, ϕ, and ψ, ideally the loss term of the discriminator will force the posterior to be roughly equal to the prior.

As long as $p(\mathbf{z})$ is easy to sample from, you can match the posterior to any distribution.

1.3 Low-ish level:

The loss of a traditional variational autoencoder is:

$$\text{ELBO} := \underbrace{\mathbb{E}_{q_\phi(z|x_i)}\Bigg[\log p_\theta(\mathbf{x_i}|\mathbf{z})\Bigg]}_{\text{①Reconstruction likelihood}} - \underbrace{D_{\text{KL}}\Bigg[q_\phi(\mathbf{z}|\mathbf{x_i})||p(\mathbf{z})\Bigg]}_{\text{②Prior constraint}} \tag{1}$$

The choice of prior constraint, however, is mostly arbitrary. As long as we have a loss term that forces $q(\mathbf{z}|\mathbf{x})$ to be near $p(\mathbf{z})$, we can use almost anything. KLD is a very natural choice and directly stems from probability distributions, but is not required. Merely replacing the KLD with the loss of the discriminator, or a similar function, results in the AAE objective.

The base paper itself was quite simple; they also presented extensions to the AAE, such as supervised AAEs, semi-supervised AAEs, clustering with AAEs, and representation learning with AAEs.

2 Empirical results:

When compared to a vanilla VAE, AAEs seem to have the following positives:

1. More distinct clustering of low dimensional images (dim 784) in low dimension latent space (dim 2 through 10).

2. Significantly better performance on semi-supervised tasks with low dimension images (dim 784).

But, they also have the following major negative:

1. Higher computational complexity, in addition to a new “link” that can fail.