Paper summary: From variational to deterministic autoencoders

May 30, 2022

1 Idea in few sentances

2 Explanation of the central concept

3 Methodology

4 Initial rambly notes

4.1 Abstract

Learning VAEs from data poses unanswered theoretical questions and considerable practical challenges. This work proposes a generative model that is simpler, deterministic, easier to train, while retaining some VAE advantages. Namely, the observation is that sampling a stochastic encoder in Gaussian VAE can be interpreted as injecting noise into the input of a deterministic decoder. The authors examine this and other regularization schemes to give smooth and meaningful latent space without forcing it to conform to an arbitrarily chosen prior.

4.2 Introduction

VAEs are great, but you have to choose between sample quality and reconstruction quality. In practise, this is attributed to overly simplistic priors or to the inherent over-regularization induced by the KL term in the VAE objective. The objective and the setup itself is problematic, read chunk of text in Introduction to see what.

The paper's contributions are the following:

1. introducing the regularized autoencoder (RAE) framework as a drop-in replacement for many common VAE archs
2. proposing an ex-post density estimation scheme which improves sample quality for (some letter)AEs without the need to retrin models
3. conducting rigorous empirical evaluation

4.3 Method

We work in VAE setting. E_ϕ is the encoder, D_θ is the decoder. The encoder deterministically maps a data point $\boldsymbol{x}$ to the mean $\mu_\phi(\boldsymbol{x})$ and variance $\sigma_\phi(\boldsymbol{x})$ in the latent space. The input to D_θ is then the mean $\mu_\phi(\boldsymbol{x})$ augmented with Gaussian noise scaled by $\sigma_\phi(\boldsymbol{x})$ via the reparametrizing trick. Authors argue that this noise injection is a key factor in having a regularized decoder (noise injection as a mean to regularize neural networks is a well-known technique). Thus training the RAE involves minimizing the simplified loss:

$$\mathcal{L}_{\text{RAE}} = \mathcal{L}_{\text{REC}} + \beta\mathcal{L}_{\boldsymbol{z}}^{\text{RAE}} + \lambda\mathcal{L}_{\text{REG}} \tag{1}$$

where $\mathcal{L}_{\text{REG}}$ represents the explicit regularizer for D_θ, and $\mathcal{L}_{\boldsymbol{z}}^{\text{RAE}} = \frac{1}{2}||\boldsymbol{z}||_2^2$, which is equivalent to contraining the size of the learned latent space, which is needed to prevent unbounded optimization. One option for $\mathcal{L}_{\text{REG}}$ is Tikhonov regularization since it is known to be related to the addition of low-magnitude input noise. In this framework this equates to $\mathcal{L}_{\text{REG}} = \mathcal{L}_{L_2} = ||\theta||_2^2$. There's also the **gradient penalty** and **spectral normalization**.

4.4 Other stuff