Notes from Understanding Variational Autoencoders

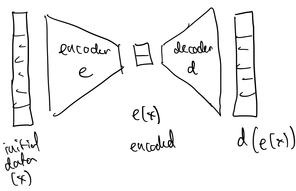
Saturday, August 27, 2022 7:07 PM

VAE - autoencoder whose encodings distribution is regularised during training to ensure latent space has good properties

notation for R.V. z, $p(z)$ is distribution of R.V.

Dimensionality reduction, PCA, and autoencoders

- dimensionality reduction - process of reducing the number of features that describe some data (through selection or extraction)



dimensionality reduction problem

find the e and d that minimize reconstruction error

$$(e^*, d^*) \quad \underset{(e,d)\in E\times D}{\operatorname{argmin}} \; \epsilon(x, d(e(x)))$$

$\epsilon(x, d(e(x)))$ → reconstruction error

- issues w/ regular autoencoders
  1. - lack of interpretable latent space
  2. - need to keep major part of data structure information in reduced representations

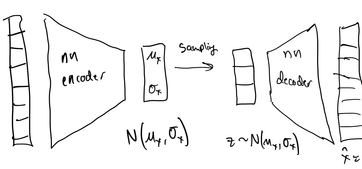
Generation - sample latent space and pass it to the decoder
- difficult to guarantee w/ traditional autoencoders
- no guarantee that model is actually learning meaningful representations

VAE: instead of encoding inputs as a single point, encode a distribution over the latent space

normal autoencoders	input x	→ encoding →	latent representation $z = e(x)$	→ decoding →	input reconstruction $d(z)$	
variational autoencoders	input x	→ encoding →	latent distribution $p(z\|x)$	→ sampling →	$z \sim p(z\|x)$ (sampled latent representation)	→ $d(z)$ input reconstruction

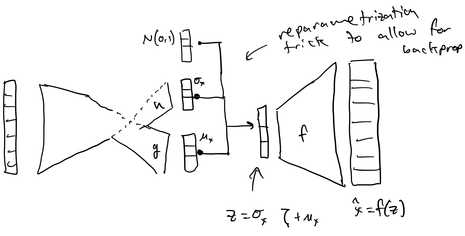
encoded distributions are chosen to be normal so encoder returns mean and covariance matrix describing distribution

VAE loss function



$$\text{loss} = \|x - \hat{x}\|^2 + KL\left[N(\mu_x, \sigma_x), N(0,1)\right]$$

reconstruction loss — regularization loss — standard normal distribution



$$\text{loss} = C\|x - f(z)\|^2 + KL\left[N(g(\tau), u(\tau)), N(0,1)\right]$$