Variational Autoencoders

Foundations of Data Analysis

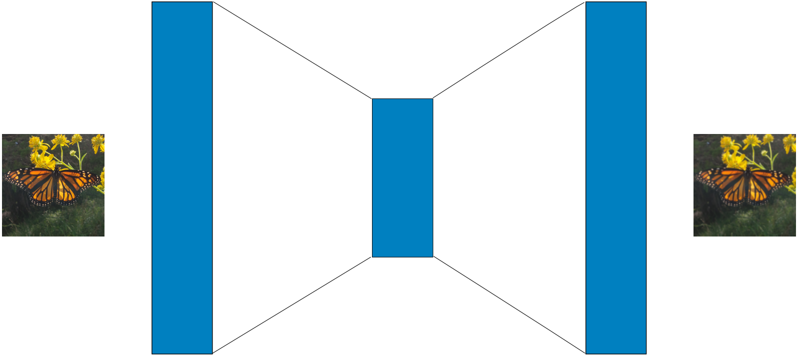
April 28, 2020

Talking about this paper:

Diederik Kingma and Max Welling, Auto-Encoding Variational Bayes, In *International Conference on Learning Representation (ICLR)*, 2014.

Autoencoders

Input | Latent Space | Output



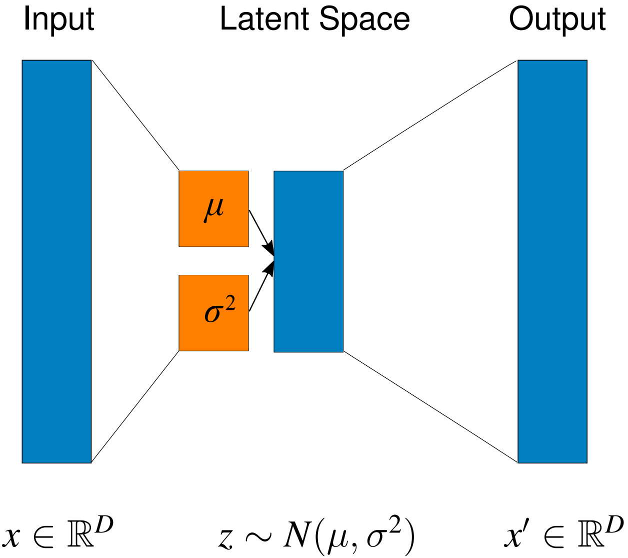
$x \in \mathbb{R}^D$ | $z \in \mathbb{R}^d$ | $x' \in \mathbb{R}^D$

$$d << D$$

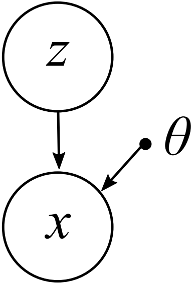
Autoencoders

- Linear activation functions give you PCA
- Training:
  1. Given data x, feedforward to x' output
  2. Compute loss, e.g., $L(x, x') = \|x - x'\|^2$
  3. Backpropagate loss gradient to update weights
- **Not** a generative model!

Variational Autoencoders



Generative Models



Sample a new x in two steps:

Prior: $p(z)$

Generator: $p_\theta(x \mid z)$

Now the analogy to the "encoder" is:

Posterior: $p(z \mid x)$

Posterior Inference

Posterior via Bayes' Rule:

$$p(z \mid x) = \frac{p_\theta(x \mid z)p(z)}{\int p_\theta(x \mid z)p(z)dz}$$

Integral in denominator is (usually) intractable!

Could use Monte Carlo to approximate, but it's expensive

Kullback-Leibler Divergence

$$\begin{aligned} D_{\mathrm{KL}}(q \| p) &= -\int q(z) \log \left(\frac{p(z)}{q(z)} \right) dz \\ &= E_q \left[-\log \left(\frac{p}{q} \right) \right] \end{aligned}$$

The average *information gained* from moving from q to p

Variational Inference

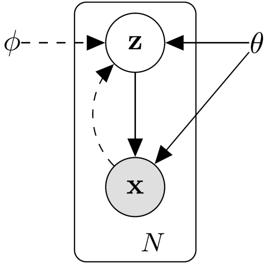
Approximate intractable posterior $p(z \mid x)$ with a manageable distribution $q(z)$

Minimize the KL divergence: $D_{\mathrm{KL}}(q(z) \| p(z \mid x))$

Variational Autoencoder

$q_\phi(z \mid x)$

Encoder Network



$p_\theta(x \mid z)$

Decoder Network

Maximize ELBO:

$$\mathcal{L}(\theta, \phi, x) = E_{q_\phi}[\log p_\theta(x, z) - \log q_\phi(z \mid x)]$$

Results from Kingma & Welling