Neural Networks and Deep Learning

Generative Models I - Variational Autoencoders

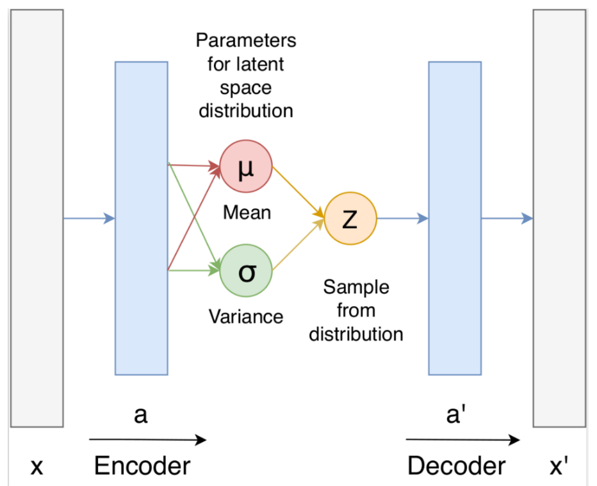
Nicholas Dronen

Department of Computer Science
dronen@colorado.edu

February 28, 2019

University of Colorado Boulder

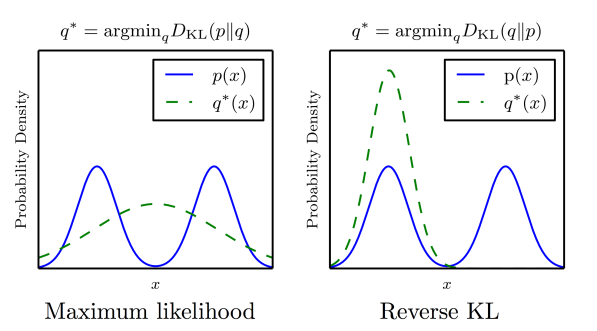
- VAE's are generative models that combine neural networks with graphical models.
- A latent variable z describes an observation x.
- To generate samples of $p(x)$, we want to sample from $p(z)$.



VAE schematic - is this plausible?

Kullback–Leibler divergence is a measure of how one probability distribution differs from a second, reference distribution.

$$D_{KL}(P||Q) = -\sum_{x \in X} P(x) log \frac{Q(x)}{P(x)}$$



Source: Goodfellow, 2016

In order to infer characteristics of z using the visible variable x, we need to find $p(z|x)$.

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$
$$p(x) = \int p(x|z)p(z)\,dz$$

p(x) is computationally intractable, however, we can apply variational inference to estimate this value.

$p(z|x)$ is estimated using another tractable distribution $q(z)$.

If we can find a $q(z)$ that is very similar to $p(z|x)$, we can use it to perform *approximate* inference of the intractable distribution.
To ensure that $q(z)$ is similar to $p(z|x)$, we minimize the KL divergence between the two distributions.

$$\min KL\left(q\left(z\right)||p\left(z|x\right)\right)$$

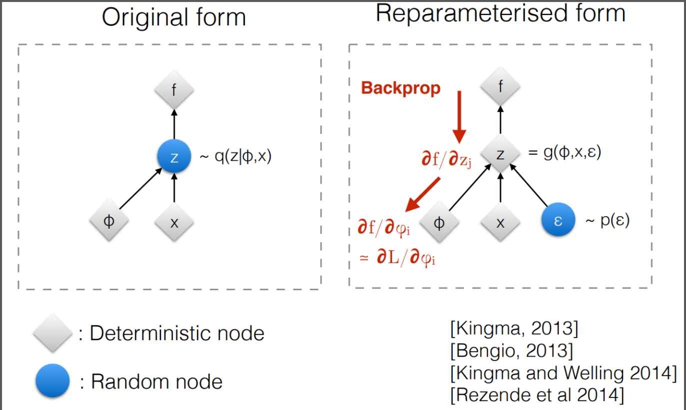
Minimizing this KL divergence is equivalent to maximizing $\mathscr{L}$, the variational lower bound.

$$\mathscr{L} = E_{q(z)} \log p\left(x|z\right) - KL\left(q\left(z\right)||p\left(z\right)\right)$$

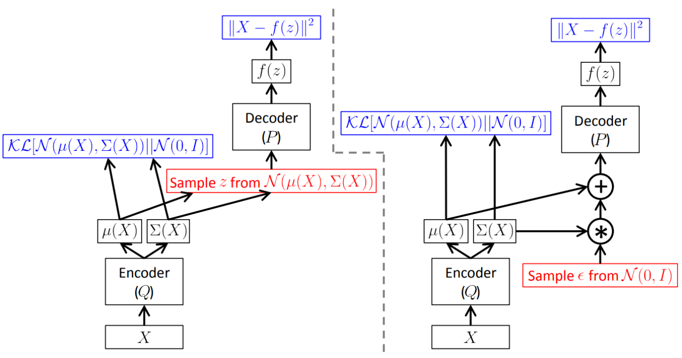
The first term represents the reconstruction likelihood and the second term ensures that our learned distribution q is similar to the true prior distribution p.

The loss function for VAEs consists of a term which penalizes reconstruction error and a term which encourages learned distribution q to be similar to true distribution p.

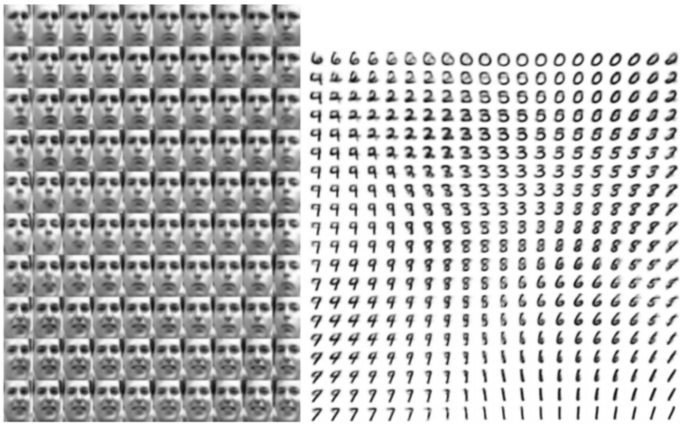
The reparameterization trick allows backpropagatation through (or around) samples from a random distribution.



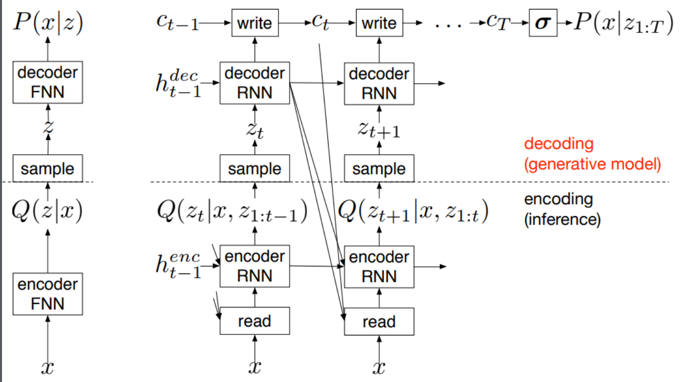
Source: Kingma, 2015



A training-time variational autoencoder implemented as a feedforward neural network, where $P(X|z)$ is Gaussian. Left is without the "reparameterization trick", and right is with it. Red shows sampling operations that are non-differentiable. Blue shows loss layers. The feedforward behavior of these networks is identical, but backpropagation can be applied only to the right network. Source: Doersch, 2016



Examples of 2-D coordinate systems for high-dimensional manifolds, learned by a variational autoencoder. Source: Kingma and Welling, 2013



Left: Conventional Variational Auto-Encoder, Right: DRAW Network.
Source: Gregor et al, 2015