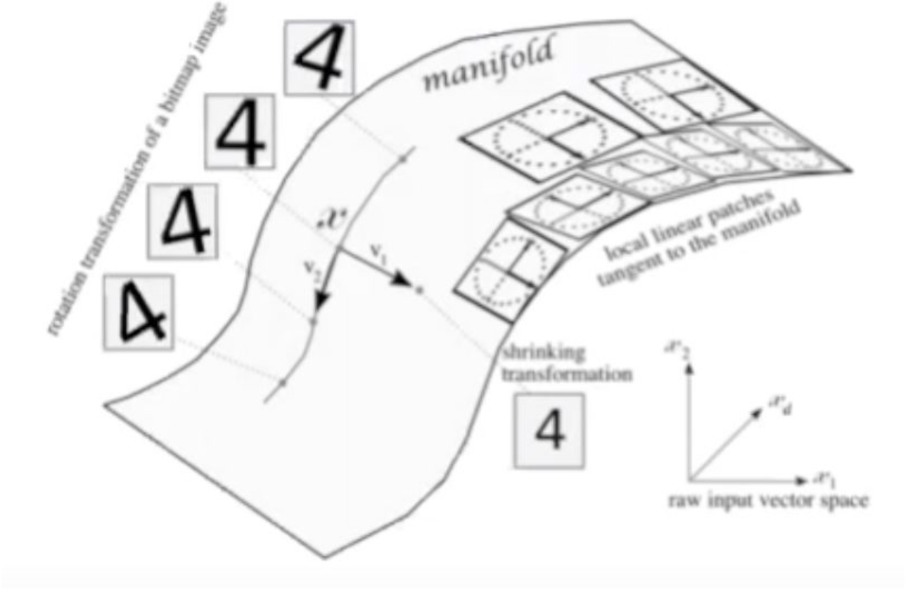
Variational AutoEncoder

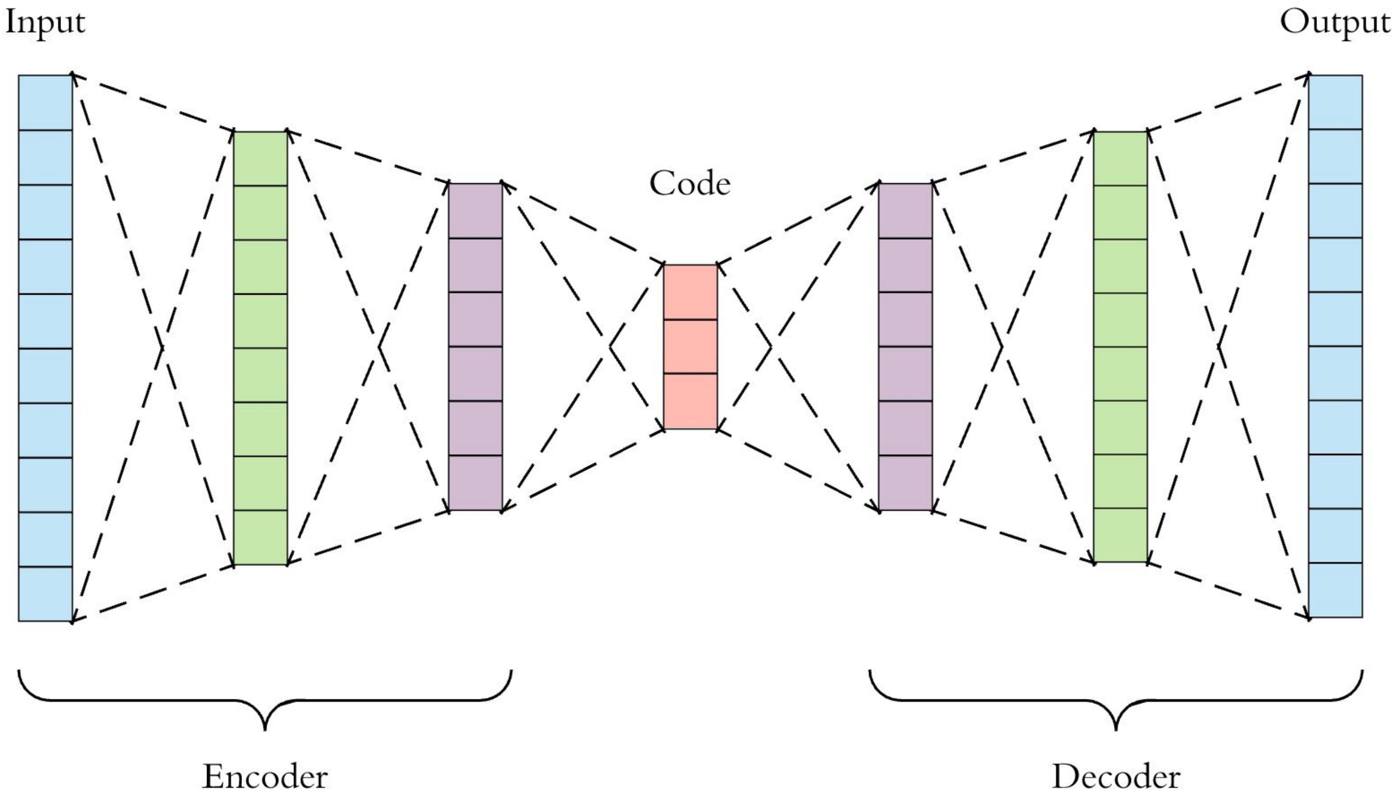
Minjae Kim

Background (Manifold Hypothesis)

- Data is concentrated around a lower dimensional manifold
- Hope finding a representation Z of that manifold from X which is high dimensional vector

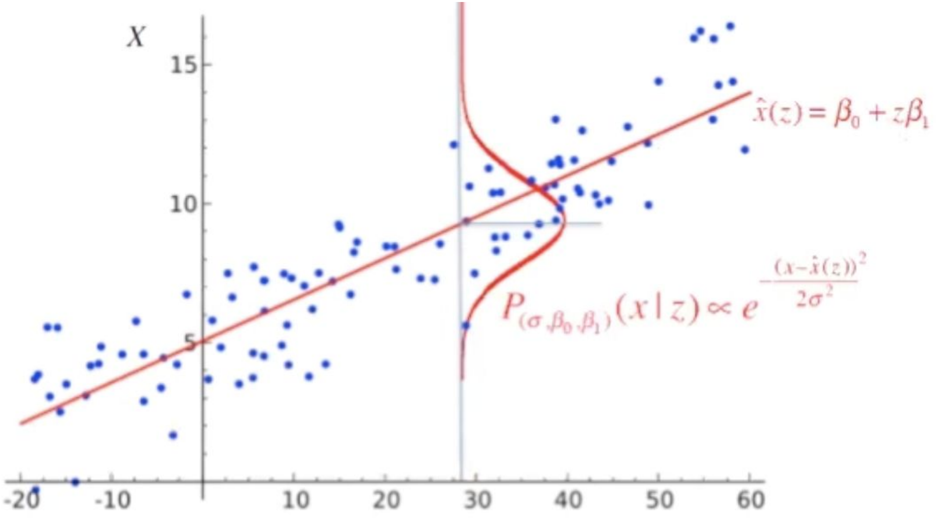


Background (Autoencoder)



Background (Linear Regression)

- We can see the linear regression problem as bayesian inference.
- Hope maximizing P(X|Z)



Background (Maximum Likelihood)

A method of estimating the parameters of a distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable.

Note that p(x) is intractable.

$$\arg\max_{\theta}\left[p_\theta(x) = \int_z p_\theta(x, z) = \int_z p_\theta(x|z)p_\theta(z)\right]$$

Variational Inference

General family of methods for approximating complicated densities by a simpler class of densities.

$$q_\phi(z|x) \approx p_\theta(z|x)$$

The variational Bound (ELBO)

We want to maximize p(x), by maximizing log p(x)

$$
\begin{aligned}
\log(p_\theta(x)) &= \int_z q_\phi(z|x)\log(p_\theta(x)) \\
&= \int_z q_\phi(z|x)\log\frac{p_\theta(z,x)}{p_\theta(z|x)} \\
&= \int_z q_\phi(z|x)\log\left(\frac{p_\theta(z,x)}{q_\phi(z|x)}\frac{q_\phi(z|x)}{p_\theta(z|x)}\right) \\
&= \int_z q_\phi(z|x)\log\frac{p_\theta(z,x)}{q_\phi(z|x)} + \int_z q_\phi(z|x)\log\frac{q_\phi(z|x)}{p_\theta(z|x)} \\
&= \mathcal{L}(\theta,\phi;x) + D_{KL}(q_\phi(z|x)||p_\theta(z|x)) \\
&\geq \mathcal{L}(\theta,\phi;x)
\end{aligned}
$$

$$\int_z q(z|x) = 1$$

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

$$\mathbb{E}[X] = \int x f(x)$$

$$D_{KL}(P||Q) = \int p(x)\log\frac{p(x)}{q(x)}$$

Because of existence of lower bound, we can maximize lower bound.

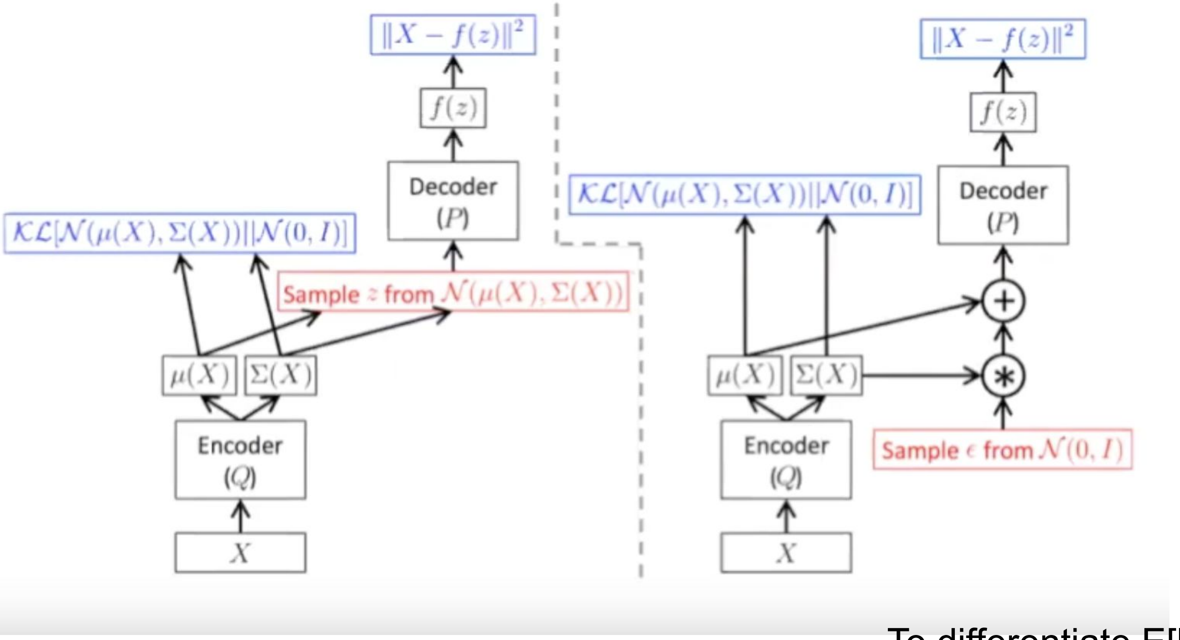
ELBO

$$\begin{aligned}\mathcal{L}(\theta,\phi;x) &= \int_z q_\phi(z|x)\log\frac{p_\theta(z,x)}{q_\phi(z|x)}\\ &= \int_z q_\phi(z|x)\log\frac{p_\theta(z)p_\theta(x|z)}{q_\phi(z|x)}\\ &= -\int_z q_\phi(z|x)\log\frac{q_\phi(z|x)}{p_\theta(z)} + \int_z q_\phi(z|x)\log p_\theta(x|z)\\ &= -D_{KL}\left(q_\phi(z|x)||p_\theta(z)\right) + \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]\end{aligned}$$

$$-D_{KL}(q(z|x)||p(z)) = \frac{1}{2}\sum_{j=1}^{J}\left(1+\log((\sigma_j)^2) - (\mu_j)^2 - (\sigma_j)^2\right)$$

We can reduce the form of ELBO like this.
The first term is Regularization term => fitting q(z|x) to p(z)
The second term is Reconstruction term => x -> q -> z -> p -> x
Where p(z) = N(0, I), q(z|x) is gaussian, D_KL can be a function of mu and sigma

Reparametrization Tricks



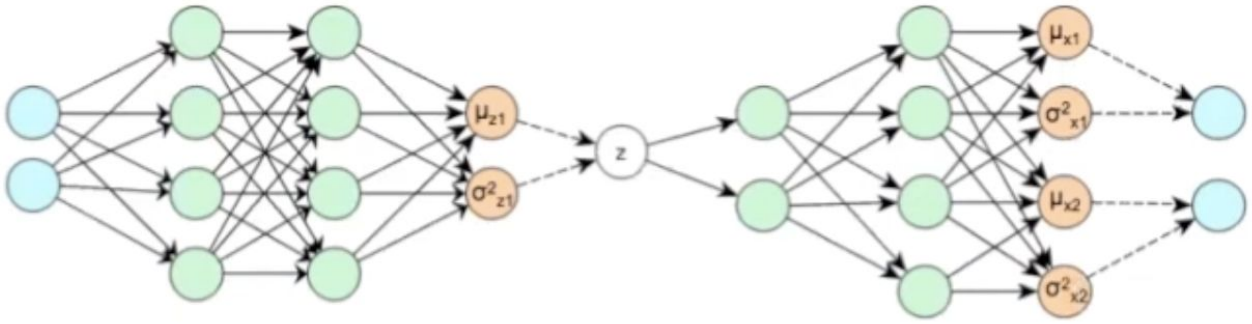
$$z = \mu_q + \sigma_q \cdot \epsilon \quad \text{where} \quad \epsilon \sim N(0,1)$$

$$\begin{aligned} & \frac{\partial}{\partial W} \mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x|z)\right] \\ = & \frac{\partial}{\partial W} \mathbb{E}_{\epsilon \sim N(0,1)} \left[\log p_\theta(x|\mu_q + \sigma_q \cdot \epsilon)\right] \\ = & \mathbb{E}_{\epsilon \sim N(0,1)} \left[\frac{\partial}{\partial W} \log p_\theta(x|\mu_q + \sigma_q \cdot \epsilon)\right] \end{aligned}$$

To differentiate E[log p(x|z)] when backpropagate, make z differentiable.

Overall

Prior p(z) ~ N(0,1) and p, q Gaussian, extension to dim(z) > 1 trivial



Cost: Regularisation

$$-D_{KL}\left(q(z|x^{(i)})\|p(z)\right) = \frac{1}{2}\sum_{j=1}^{J}\left(1 + \log(\sigma_{z_j}^{(i)^2}) - \mu_{z_j}^{(i)^2} - \sigma_{z_j}^{(i)^2}\right)$$

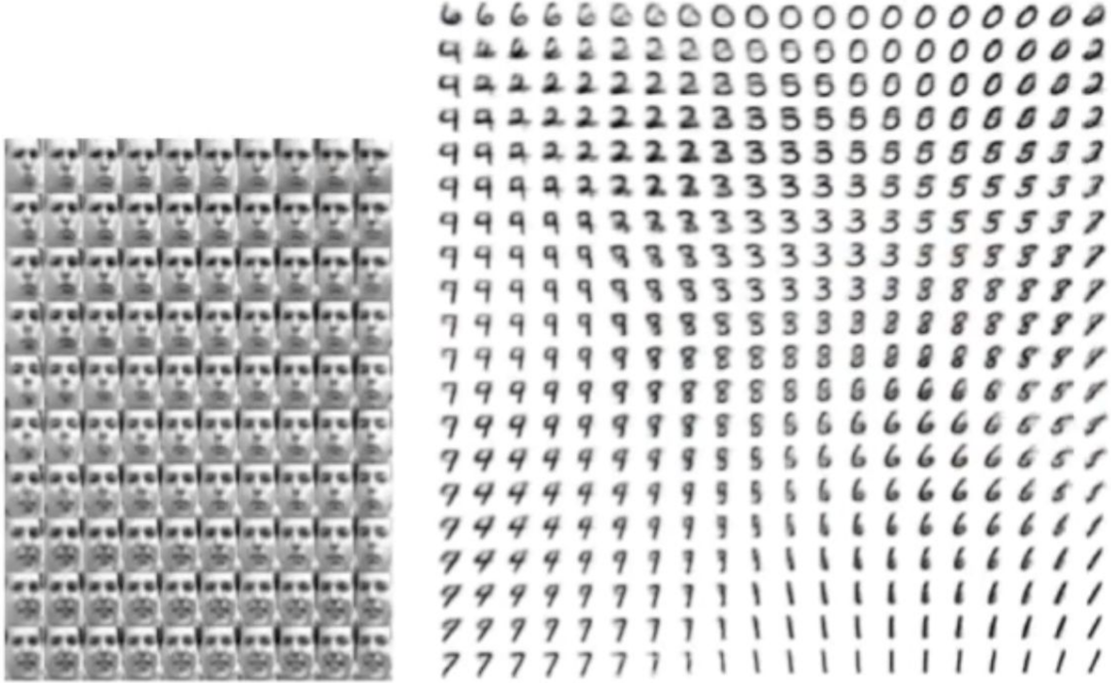
Cost: Reproduction

$$-\log\left(p(x^{(i)}|z^{(i)})\right) = \sum_{j=1}^{D}\frac{1}{2}\log(\sigma_{x_j}^2) + \frac{(x_j^{(i)} - \mu_{x_j})^2}{2\sigma_{x_j}^2}$$

We use mini batch gradient decent to optimize the cost function over all $x^{(i)}$ in the mini batch

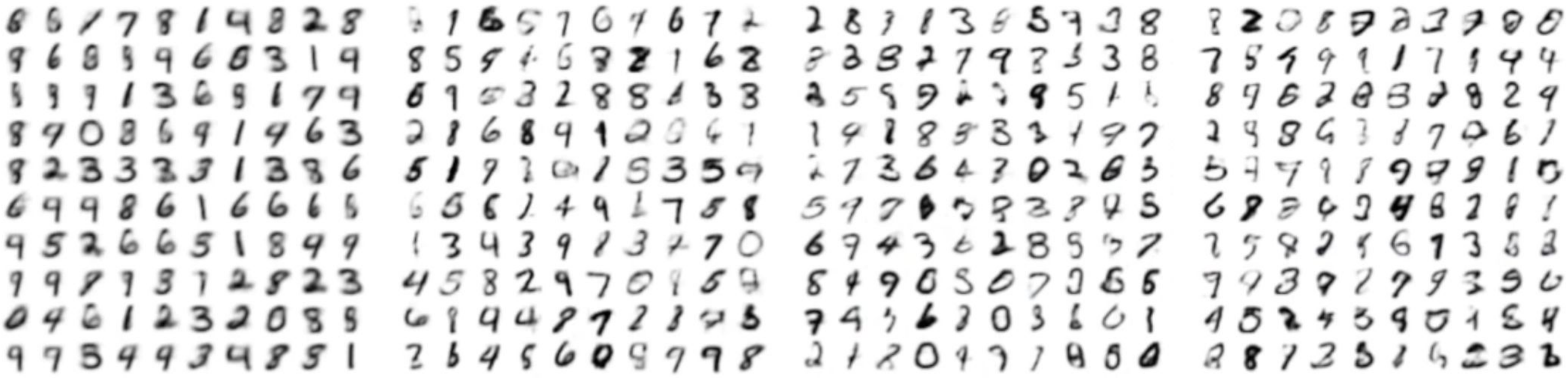
Least Square for constant variance

Result



(a) Learned Frey Face manifold (b) Learned MNIST manifold

Result (cont.)



(a) 2-D latent space (b) 5-D latent space (c) 10-D latent space (d) 20-D latent space