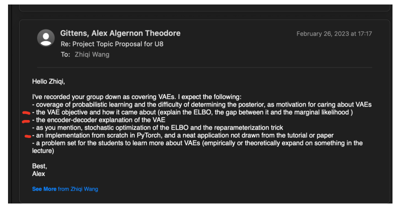
Motivation : Question : what's log likelihood


Gittens, Alex Algernon Theodore
February 26, 2023 at 17:17
Re: Project Topic Proposal for U8
To: Zhiqi Wang

Hello Zhiqi,

I've recorded your group down as covering VAEs. I expect the following:
- coverage of probabilistic learning and the difficulty of determining the posterior, as motivation for caring about VAEs
- the VAE objective and how it came about (explain the ELBO, the gap between it and the marginal likelihood)
- the encoder-decoder explanation of the VAE
- as you mention, stochastic optimization of the ELBO and the reparameterization trick
- an implementation from scratch in PyTorch, and a neat application not drawn from the tutorial or paper
- a problem set for the students to learn more about VAEs (empirically or theoretically expand on something in the lecture)

Best,
Alex

See More from Zhiqi Wang

Going from AE (lecture 17)

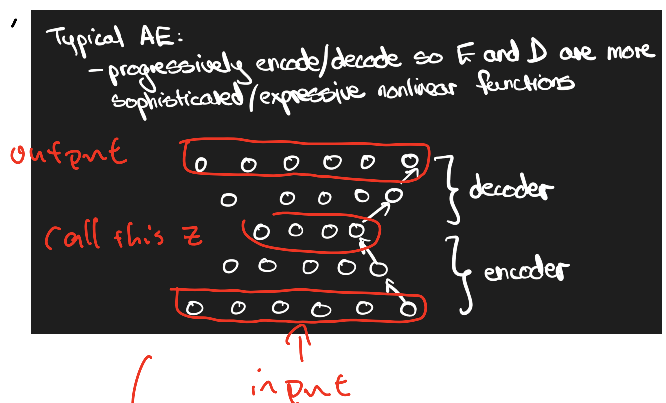
AE is a Deep Learning model that encode high dimensional information into latent space (encoder) to learn a lower dimension representation while ignoring the noise.

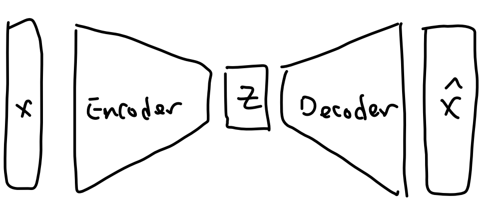
In assignment 4, we also use a decoder to reconstruct an input image, but it's just an reconstruction.

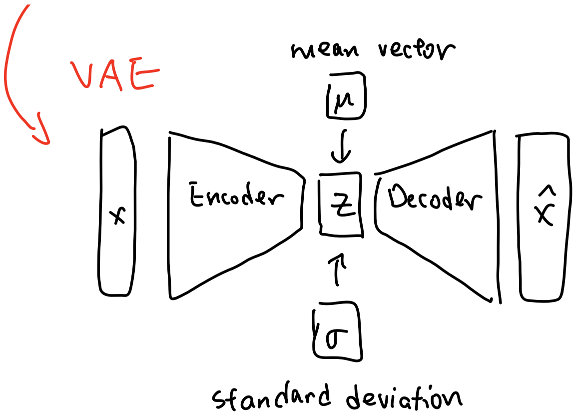
Think about latent space representation as features of an image, how to generate new image (not in the train set) from given features ?

VAE (variational autoencoder)

Let's take prof. Gittens'







Encoder : Inference model

Decoder : Generative model

The encoder computes $p_\phi(z|x)$

The decoder computes $q_\theta(x|z)$

And the $\boxed{\mu}$ and $\boxed{\sigma}$ comes from a fixed prior on the latent distribution. A common choice of prior is Normal distribution.

ELBO : Evidence Lower Bound

ϕ : variational parameters

Marginal Likelihood

$$\frac{q_\theta(x,z)}{q_\theta(z|x)} = \frac{\cancel{q_\theta(x,z)}}{\frac{\cancel{q_\theta(z,x)}}{q_\theta(x)}} = q_\theta(x)$$

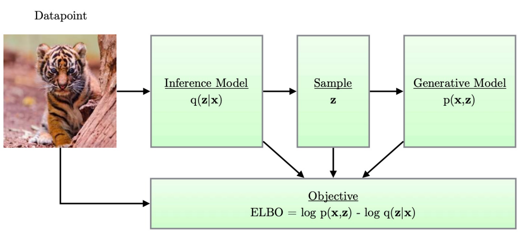
$$\begin{aligned}
\log p_\theta(x) &= \mathbb{E}_{p_\phi(z|x)}[\log q_\theta(x)] \\
&= \mathbb{E}_{p_\phi(z|x)}\left[\log\left[\frac{q_\theta(x,z)}{q_\theta(z|x)}\right]\right] \\
&= \mathbb{E}_{p_\phi(z|x)}\left[\log\left[\frac{q_\theta(x,z)}{p_\phi(z|x)} \frac{p_\phi(z|x)}{q_\theta(z|x)}\right]\right] \quad \text{"chain rule?"} \\
&= \underbrace{\mathbb{E}_{p_\phi(z|x)}\left[\log\left[\frac{q_\theta(x,z)}{p_\phi(z|x)}\right]\right]}_{\substack{= L_{\theta,\phi}(x) \\ \text{(ELBO)}}} + \underbrace{\mathbb{E}_{p_\phi(z|x)}\left[\log\left[\frac{p_\phi(z|x)}{q_\theta(z|x)}\right]\right]}_{\substack{= D_{KL}(p_\phi(z|x) \| q_\phi(z|x)) \\ \text{KL Divergence}}}
\end{aligned}$$

Since D_{KL} is non-negative, therefore L_θ is the lower bound of $\log p_\theta(x)$.

$$\log p_\theta(x) = L_{\theta,\phi}(x) + D_{KL}(p_\phi(z|x) \| q_\phi(z|x))$$

$$D_{KL}(p_\phi(z|x) \| q_\phi(z|x)) = \log p_\phi(x) - L_{\theta,\phi}(x)$$

Therefore, D_{KL} is often called the gap between marginal likelihood and ELBO. Which being how well $p_\phi(z|x)$ approximates the true (posterior) distri. $q_\theta(z|x)$ in terms of D_{KL}.



from An Introduction to Variational Autoencoders, I used different notation

Objective:

$$\phi^* = \underset{\phi}{\operatorname{argmax}}\ ELBO$$

$$= \underset{\phi}{\operatorname{argmax}} \log p_\theta(x) - \underset{\phi}{\operatorname{argmin}}\ D_{KL}(p_\phi(z|x) \| q_\phi(z|x))$$

- $\underset{\phi}{\operatorname{argmax}} \log p_\theta(x)$: maximizing the marginal likelihood, means our generative model gets better.

- $\underset{\phi}{\operatorname{argmin}} D_{KL}(p_\phi(z|x) || q_\phi(z|x))$: minimizing KL Divergence of the approximation from the true posterior, means our inference model gets better.

SGD & Reparameterization Trick