Reparameterization trick

Assume $q_\phi(\mathbf{z}|\mathbf{x}^i) = \prod_k q_\phi(z_k|\mathbf{x}^i)$

Let $q_\phi(z_k|\mathbf{x}^i) \sim \mathcal{N}(\mu_{z_k|\mathbf{x}^i}, \sigma^2_{z_k|\mathbf{x}^i})$

To sample from $\mathcal{N}(\mu_{\mathbf{z}|\mathbf{x}^i}, \sigma^2_{\mathbf{z}|\mathbf{x}^i})$, we use **Reparameterization trick**.

Sample from a parameterless distribution. Here, we choose $\mathcal{N}(0, I)$.

Let the samples be $v_1 \dots v_r$

$z_j = v_j \sigma_{z|\mathbf{x}^i} + \mu_{z|\mathbf{x}^i}$ are samples from $\mathcal{N}(\mu_{z|\mathbf{x}^i}, \sigma_{z|\mathbf{x}^i})$.

$\mu_{z|x^i}, \sigma_{z|x^i} \sim q_\phi(x^i)$

$z_j \sim \mathcal{N}(\mu_{z_j|x}, \sigma^2_{z_j|x})$

$\nabla_\phi z_j = v_j \nabla_\phi \sigma_{z|x^i} + \nabla_\phi \mu_{z|x^i}$

$v_j \sim \mathcal{N}(0, 1)$

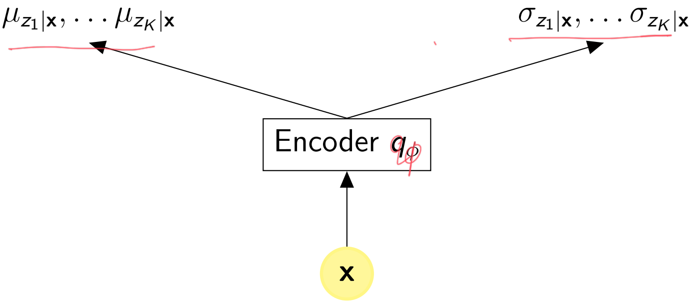
$v_j = \dfrac{z_j - \mu_{z_j|x}}{\sigma_{z_j|x}} \sim \mathcal{N}(0, 1)$

Training VAE(Continued)

depend on ϕ

$$\max_{\theta,\phi} \sum_{i=1}^{N} \frac{1}{r} \sum_{j=1}^{r} \log P_\theta(\mathbf{x}^i|\mathbf{z}_j) = \max_{\theta,\phi} \sum_{i=1}^{N} \frac{1}{r} \sum_{j=1}^{r} \log P_\theta(\mathbf{x}^i|\mu_{\mathbf{z}|\mathbf{x}^i} + \sigma_{\mathbf{z}|\mathbf{x}^i} * \mathbf{v}_j)$$

v_j samples do not involve ϕ $\sim \mathcal{N}(0, I)$



Calculating second term: KL distance

$\int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}^i) \log \frac{q_\phi(\mathbf{z}|\mathbf{x}^i)}{P_\theta(z)})$ is the KL distance between $q_\phi(\mathbf{z}|\mathbf{x}^i)$ and $\boxed{P_\theta(z)}$.

$z \in \{z_1, z_2 \cdots z_K\}$

KL distance in closed form

Assume $P_\theta(z) = \mathcal{N}(0,1)$, $q(z|x^i) = \mathcal{N}(\mu_i, \sigma_i^2)$

$KL(q(x); p(x))$

$= \int_x q(x) \log \frac{q(x)}{p(x)}$

$KL(q(x); q(x)) = 0$

$\mu_i \equiv \mu_{z_k|x^i}$

$$\nabla_\phi KL(q(z|x), P(z)) = \int_z q(z|x) \log \frac{q(z|x^i)}{P(z)}$$

$$= \int_z \mathcal{N}(\mu_i, \sigma_i^2) 0.5[-\frac{(z-\mu_i)^2}{\sigma_i^2} - \log \sigma_i^2 + z^2]$$

$$= 0.5[-\frac{\sigma_i^2}{\sigma_i^2} - \log \sigma_i^2 + \int_z z^2 \mathcal{N}(\mu_i, \sigma_i^2)]$$

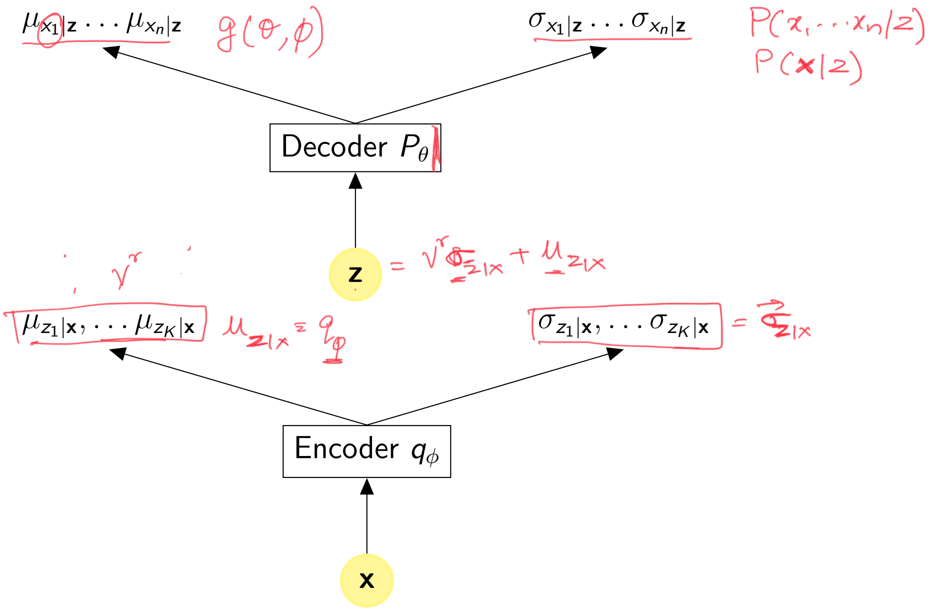
$$\boxed{\nabla_\phi = 0.5[-1.0 - \log \sigma_i^2 + \mu_i^2 + \sigma_i^2]}$$

2nd part of training objective.

Writing in full notation:

$$\sum_{i=1}^{B} \sum_{k=1}^{K} 0.5[-1.0 - \log \sigma^2_{z_k|x^i} + \mu^2_{z_k|x^i} + \sigma^2_{z_k|x^i}]$$

Putting it all together (Training VAEs)



Overall Training Algorithm

Initialize θ and ϕ network parameters randomly.
for number of training iterations **do**

Sample minibatch of B examples $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^B$ from data D

Compute $\mu_{\mathbf{z}|\mathbf{x}^i}, \sigma_{\mathbf{z}|\mathbf{x}^i} \leftarrow q_\phi(\mathbf{x}^i)$ $\qquad N(0, I)$

Get Br samples $\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^{Br}$ from noise prior $P(\mathbf{z})$. Each is of K dims.

Compute $\boxed{\mathbf{z}^{i,r}} = \sigma_{\mathbf{z}|\mathbf{x}^i}\mathbf{v}^{i,r} + \mu_{\mathbf{z}|\mathbf{x}^i}$

Compute $\boxed{\mu_{\mathbf{x}^i|\mathbf{z}}, \sigma_{\mathbf{x}^i|\mathbf{z}}} \leftarrow P_\theta(\mathbf{z}^{i,r})$ Decoder.

$\min_{\theta,\phi} \sum_{i,r} \frac{1}{r} \log N(\mathbf{x}^i|\mu_{\mathbf{x}^i|\mathbf{z}}, \sigma_{\mathbf{x}^i|\mathbf{z}}) + \boxed{\sum_i \mu^2_{\mathbf{x}^i|\mathbf{z}} + \sigma^2_{\mathbf{x}^i|\mathbf{z}} - \log \sigma_{\mathbf{x}^i|\mathbf{z}}}$

$P(x^i | \sigma_{x|z^i}, \mu_{x|z^i})$ $\qquad$ $z|x^i$ KL-distance.

end for

$$\nabla_\theta \left(\sum_i \frac{(x^i - \mu_{x|z^{i,r}})^2}{2\sigma^2_{x|z}} + \log \sigma_{x|z^{i,r}} + \underbrace{\sum_i \mu^2_{z|x^i} + \sigma^2_{z|x^i} - \log \sigma_{z|x^i}}_{\text{constant wrt } \theta} \right)$$

LL(θ,z) depend on θ

$z^{i,r} \sim P(z|x^i) \approx q_\phi(z|x^i)$

$\nabla_\phi = \nabla LL_z \nabla_\phi z$

$\mu_{x|z^{i,r}} = P(x|z^{i,r}) \sim N(\ , \sigma^2_{x|z^{i,r}})$

Inference on Trained VAEs

1. Generating new samples **x**

Sample $z \sim P(z)$ which is often $\mathcal{N}(0, I)$

compute $\mu_{x|z}$ & $\sigma_{x|z}$

Sample $x \sim \mathcal{N}(\mu_{x|z}, \sigma^2_{x|z})$.

2. Finding latent factors of a given **x** $\quad P(z|x)$

$x \longrightarrow q_\phi \rightarrow \mu_{z|x} \quad P(z|x) \sim \mathcal{N}(\mu_{z|x}, \sigma^2_{z|x})$

$q_\phi \rightarrow \sigma_{z|x}$

Additional Resources

- Code Sample: https://github.com/hwalsuklee/tensorflow-mnist-VAE
- Unsupervised Deep Learning, NeurIPS 2018 Tutorial, https://nips.cc/Conferences/2018/Schedule?showEvent=10985