Variational Autoencoder

Agenda

- VAE
  - Models with Latent Variables
  - Probabilistic PCA
  - Amortized Variational Inference
  - Reparametrization Trick

Probabilistic PCA

Consider that data $x \in \mathbb{R}^D$ and we want to get the low-dimensional representation $z \in \mathbb{R}^d$. The model has form

$$x = \mu + Wz + \varepsilon, \tag{1}$$

where $\mu \in \mathbb{R}^D,\ W \in \mathbb{R}^{D\times d},\ z \in \mathbb{R}^d$, and $\varepsilon \in \mathbb{R}^D$.
If we assume that ε has normal distribution with variance σ^2, then the model with latent variable takes form:

$$p(x|z,\ \theta) = p_\theta(x|z) = N(x|\mu + Wz,\ \sigma^2 I), \tag{2}$$

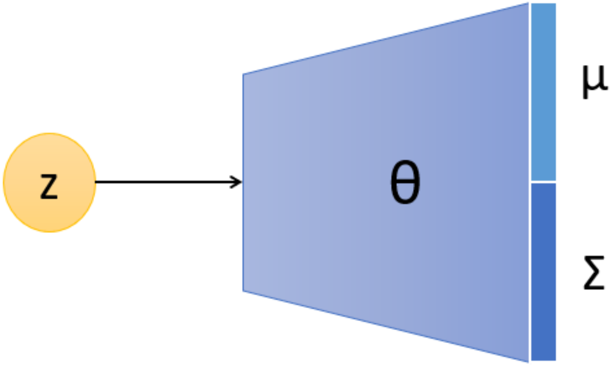
with parameters $\theta = \{\mu, W, \sigma)\}$.
If we assume that the prior distribution is a standard Gaussian, then the joint probability can be written as follows:

$$p_\theta(x,\ z) = p_\theta(x|z)p(z) = N(x|\mu + Wz,\ \sigma^2 I)N(t|0,\ I). \tag{3}$$

VAE (Generator)

$$p_\theta(x,\ z) = p_\theta(x|z)p(z) = N(x|\mu + Wz,\ \sigma^2 I)N(t|0,\ I).$$

$$p_\theta(x,\ z) = p_\theta(x|z)p(z) = N(x|\mu(z, \theta),\ \Sigma(z, \theta))N(t|0,\ I).$$



Partial Likelihood Estimation

$$L(\theta) = \log p_\theta(X) = \log \prod_{i=1}^{N} p_\theta(x^{(i)}) = \sum_{i=1}^{N} \log p_\theta(x^{(i)}) \to \max_\theta. \quad (4)$$

We can multiply each term by $1 = \int q^{(i)}(z)dz$, where $q^{(i)}(z)$ is an arbitrary distribution that we also denote by $q(z|x^{(i)})$:

$$L(\theta) = \sum_{i=1}^{N} \int q^{(i)}(z) \log p_\theta(x^{(i)}) dz. \quad (5)$$

Using the conditional probability formula, we obtain

$$L(\theta) = \sum_{i=1}^{N} \int q^{(i)}(z) \log \frac{p_\theta(x^{(i)},\ z)}{p_\theta(z|x^{(i)})} \frac{q^{(i)}(z)}{q^{(i)}(z)} dz. \quad (6)$$

ELBO

$$L(\theta) = \sum_{i=1}^{N} \int q^{(i)}(z) \log \frac{p_\theta(x^{(i)},\ z)}{p_\theta(z|x^{(i)})} \frac{q^{(i)}(z)}{q^{(i)}(z)} dz.$$

Splitting the logarithm, we have

$$L(\theta) = \sum_{i=1}^{N} \underbrace{\int q^{(i)}(z) \log \frac{p_\theta(x^{(i)},\ z)}{q^{(i)}(z)} dz}_{\mathcal{L}(q^{(i)}(z),\ \theta)} + \sum_{i=1}^{N} \underbrace{\int q^{(i)}(z) \log \frac{q^{(i)}(z)}{p_\theta(z|x^{(i)})} dz}_{KL(q^{(i)}(z)||p_\theta(z|x^{(i)}))}.$$

$$L(\theta) \geq \sum_{i=1}^{N} \mathcal{L}(q^{(i)}(z),\ \theta) \equiv \mathcal{L}(q(Z),\ \theta) \qquad (ELBO).$$

ELBO

$$\mathcal{L}(q(Z),\ \theta) = \sum_{i=1}^{N} \int q(z|x^{(i)}) \log \frac{p_\theta(x^{(i)},\ z)}{q(z|x^{(i)})} dz$$

$$= \sum_{i=1}^{N} \int q(z|x^{(i)}) \log p_\theta(x^{(i)}|\ z) dz + \sum_{i=1}^{N} \int q(z|x^{(i)}) \log \frac{p(z)}{q(z|x^{(i)})} dz$$

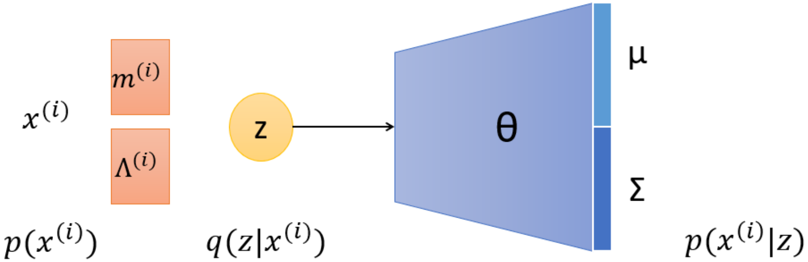
$$= \sum_{i=1}^{N} \int q(z|x^{(i)}) \log p_\theta(x^{(i)}|\ z) dz - \sum_{i=1}^{N} KL\left(q(z|x^{(i)})||p(z)\right)$$

We will search q in a family of factorized Gaussian distributions:

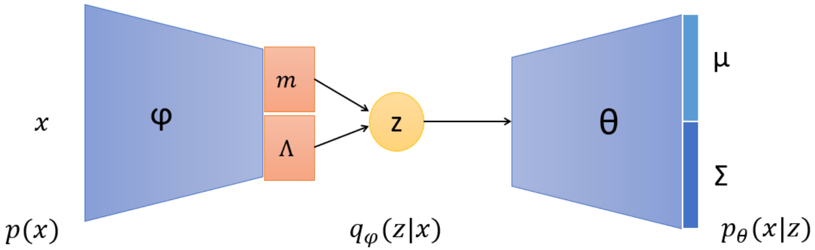
$$q(z|x^{(i)}) = N(m^{(i)}, \Lambda^{(i)}).$$

$$KL\left(q(z|x^{(i)})||p(z)\right) = \frac{1}{2}\sum_{s=1}^{d}\left(m_s^{(i)2} + \lambda_s^{(i)2} - 1 - \log \lambda_s^{(i)2}\right)$$

Amortized Variational Inference



Amortized Variational Inference



Reparametrization Trick

$$\mathcal{L}(\varphi,\ \theta) = \sum_{i=1}^{N} \int q_\varphi(z|x^{(i)}) \log p_\theta(x^{(i)}|\ z)dz - \sum_{i=1}^{N} KL\left(q_\varphi(z|x^{(i)})||p(z)\right)$$

$$\underset{\varphi}{\text{stochgrad}}\mathcal{L}(\varphi,\ \theta) = N\frac{\partial}{\partial\varphi} \int q_\varphi(z|x^{(i)}) \log p_\theta(x^{(i)}|\ z)dz - N\frac{\partial}{\partial\varphi}KL(q_\varphi(z|x^{(i)})||$$

$$\underset{\theta}{\text{stochgrad}}\mathcal{L}(\varphi,\ \theta) = N\frac{\partial}{\partial\theta} \int q_\varphi(z|x^{(i)}) \log p_\theta(x^{(i)}|\ z)dz$$

Reparametrization Trick

$$z = g(x^{(i)}, \varphi, \varepsilon) = m(x^{(i)}, \varphi) + \Lambda(x^{(i)}, \varphi)\varepsilon, \qquad \varepsilon \sim r(\varepsilon) = N(0, I).$$

$$\begin{aligned}
\mathcal{L}(\varphi,\ \theta) &= \sum_{i=1}^{N} \int q_{\varphi}(z|x^{(i)}) \log p_{\theta}(x^{(i)}|\ z)dz - \sum_{i=1}^{N} KL\left(q_{\varphi}(z|x^{(i)})||p(z)\right) \\
&= \sum_{i=1}^{N} \int r(\varepsilon) \log p_{\theta}(x^{(i)}|\ g(x^{(i)}, \varphi, \varepsilon))d\varepsilon - \sum_{i=1}^{N} KL\left(q_{\varphi}(z|x^{(i)})||p(z)\right)
\end{aligned}$$

$$\underset{\varphi}{\text{stochgrad}} \mathcal{L}(\varphi,\ \theta) = N\frac{\partial}{\partial \varphi} \log p_{\theta}(x^{(i)}|\ g(x^{(i)}, \varphi, \hat{\varepsilon})) - N\frac{\partial}{\partial \varphi} KL(q_{\varphi}(z|x^{(i)})||p(z))$$

$$\underset{\theta}{\text{stochgrad}} \mathcal{L}(\varphi,\ \theta) = N\frac{\partial}{\partial \theta} \log p_{\theta}(x^{(i)}|\ g(x^{(i)}, \varphi, \hat{\varepsilon}))$$

$$\hat{\varepsilon} \sim N(0, I).$$

Sampling