Variational Autoencoders [Kingma, Welling]

Setup: Given $\{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^d$, $d \gg 1$

Goal: Learn the distribution that generated $x \sim p(x)$.

Assumption: There exists a set of latent variables $z \in \mathbb{R}^q$, $q \ll$

explain the variability observed in x.

Model: $$P_\theta(x) = \int \underbrace{p_\theta(x,z)}_{\text{generative model}} dz = \int \underbrace{P_\theta(x|z)}_{\text{likelihood}} \underbrace{p(z)}_{\text{prior}} dz$$

It is natural to ask what those latent variables should be z?

$$\underbrace{p(z|x)}_{\text{posterior}} = \frac{P_\theta(x|z)\, p(z)}{P_\theta(x)}$$

↳ intractable

VAE model assumptions:

i.) $p(z|x) \approx q_\varphi(z|x) = \mathcal{N}(\mu_\varphi(x), \overbrace{\Sigma_\varphi(x)}^{\text{diag.}})$: encoder   variational posterior /

ii.) $P_\theta(x|z) = \mathcal{N}(\mu_\theta(z), \overbrace{\Sigma_\theta(z)}^{\text{diag.}})$: decoder   likelihood /

iii) $P_\theta(z) = \mathcal{N}(0, I)$: prior

Parameters : $\{\theta, \varphi\}$, $\theta^*, \varphi^* = \underset{\theta,\varphi}{\operatorname{argmin}} -\log p_\theta(x)$

Variational inference :

$$-\log p_\theta(x) = -\log \int p_\theta(x,z)\,dz \overset{I.S}{=} -\log \int \frac{p_\theta(x,z)}{q_\varphi(z|x)} q_\varphi($$

$$= -\log \mathbb{E}_{z\sim q_\varphi(z|x)} \left[\frac{p_\theta(x|z)p(z)}{q_\varphi(z|x)} \right]$$

Jensen

$$\leq -\mathbb{E}_{z\sim q_\varphi(z|x)} \left[\log \frac{p_\theta(x|z)p(z)}{q_\varphi(z|x)} \right]$$

$$= -\int \log \frac{\overbrace{p_\theta(x|z)}\,p(z)}{q_\varphi(z|x)} q_\varphi(z|x)\,dz$$

$$= -\int \log p_\theta(x|z)\, q_\varphi(z|x)\,dz + \overbrace{\int \log \frac{q_\varphi(z|x)}{p(z)} q_\varphi(z|x)\,(}^{\mathbb{KL}}$$

$$= -\mathbb{E}_{z\sim q_\varphi(z|x)} \left[\log p_\theta(x|z) \right] + \mathbb{KL}\left[q_\varphi(z|x) \,\|\, p(z) \right]$$

$$\Longrightarrow \underline{\mathcal{L}(\theta,\varphi)} := -\log p_\theta(x|z) \leq -\mathbb{E}_{z\sim q_\varphi(z|x)} \left[\log p_\theta(x|z) \right] + \mathbb{KL}\left[q_\varphi(z|x) \right.$$

Remark : To optize this objective (Evidence Lower Bound - ELBO

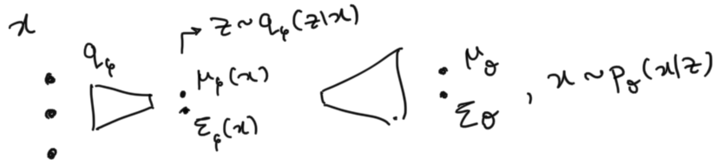
we need to compute $\nabla_\theta \mathcal{L}(\theta,\varphi)$, $\nabla_\varphi \mathcal{L}(\theta,\varphi)$.

we

$$\nabla_\varphi \mathbb{E}_{z\sim q_\varphi(z|x)}\left[\log p_\theta(x|z)\right] \longrightarrow \nabla_\varphi \mathbb{E}_{\varepsilon\sim p(\varepsilon)}\left[\log p_\theta(x \mid \underbrace{\mu_\varphi(z)+\varepsilon\Sigma}_{\text{reparam. trick}}\right.$$

for a mul

Norma



1. Input data batch $\boxed{X \xrightarrow{q_\varphi} \mu_\varphi(X), \Sigma_\varphi(X)}$

2. $z \sim q_\varphi(z|x)$: $z = \mu_\varphi(x) + \varepsilon\, \Sigma_\varphi^{-\frac{1}{2}}(x)$, $\varepsilon \sim \mathcal{N}(0, I)$

3. $\boxed{z \xrightarrow{p_\theta} \mu_\theta(z), \Sigma_\theta(z)}$, $x \sim p_\theta(x|z) = \mathcal{N}\left(\mu_\theta(z), \Sigma_\theta(z)\right)$

$$\boxed{\mathcal{L}(\theta, \varphi) := \underbrace{-\mathbb{E}_{z\sim q_\varphi(z|x)}\left[\log p_\theta(x|z)\right]}_{\text{reconstruction loss}} + \underbrace{\mathbb{KL}\left[q_\varphi(z|x) \,\|\, p(z)\right]}_{\text{regularization}}}$$

(*) Neural net architecture :

$z \longrightarrow \mu_\theta(z), \Sigma_\theta(z)$

$$z \rightarrow \boxed{\text{MLP}} \rightarrow H \begin{cases} \mu = HW_1 + b_1 \\ \Sigma = HW_2 + b_2 \end{cases}$$