Saturday, 11 March 2023 9:57 AM

Maximum Likelihood Estimation for 1-D Gaussian

$$P_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \qquad \theta = (\mu, \sigma)$$

$$\log P_\theta(x) = -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}(x-\mu)^2$$

Data : $\{x^{(i)} : i = 1, \dots, N\}$.

The log-likelihood is

$$L(x^{(1)}, \dots, x^{(N)}, \theta) = \frac{1}{N}\sum_{i=1}^{N} \log P_\theta(x^{(i)})$$

$$= \frac{1}{N}\sum_{i=1}^{N} -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}(x^{(i)}-\mu)^2$$

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{N} \frac{1}{\sigma^2}(x^{(i)}-\mu) = 0$$

$$\Rightarrow \quad \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x^{(i)} \qquad \text{sample mean}$$

$$\frac{\partial L}{\partial \sigma} = \sum_{i=1}^{N} -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x^{(i)}-\hat{\mu})^2 = 0$$

$$\Rightarrow \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)}-\hat{\mu})^2$$

Evidence Lower Bound (ELBO)

log-likelihood of latent model :

$$\log P_\theta(x) = \log \int P_\theta(x|z) P_\theta(z)\, dz$$

For some distribution $\ell_x(z)$ on z (to be set later)

$$\log p_\theta(x) = \int \log p_\theta(x) \cdot \ell_x(z)\, dz \qquad \int \ell_x(z)\, dz = 1$$

$$= \mathbb{E}_{\ell_x(z)}\left[\log p_\theta(x)\right]$$

$$= \mathbb{E}_{\ell_x(z)}\left[\log\left(\frac{p_\theta(x,z)}{p_\theta(z|x)}\right)\right] \qquad \text{Bayes Theorem}$$

$$= \mathbb{E}_{\ell_x(z)}\left[\log\left(\frac{p_\theta(x,z)}{q_\varphi(z|x)} \cdot \frac{q_\varphi(z|x)}{p_\theta(z|x)}\right)\right]$$

$$= \mathbb{E}_{\ell_x(z)} \log\left(\frac{p_\theta(x,z)}{q_\varphi(z|x)}\right) + \mathbb{E}_{\ell_x(z)} \log\left(\frac{q_\varphi(z|x)}{p_\theta(z|x)}\right)$$

Choose $\ell_x(z) = q_\varphi(z|x)$

$$\log p_\theta(x) = \underbrace{\mathbb{E}_{q_\varphi(z|x)} \log\left(\frac{p_\theta(x,z)}{q_\varphi(z|x)}\right)}_{\text{Evidence lower bound (ELBO)}} + \underbrace{D_{KL}\left(q_\varphi(z|x) \,|\, p_\theta(z|x)\right)}_{\text{KL-divergence } \geq 0}$$