Method 2 : Evidence lower Bound

Making $P(z|x)$ tractable with variational inference

$$P(z|x) = \frac{P(x,z)}{P(x)} = \frac{P(x,z)}{\int P(x,z)\,dz}$$

Variational inference approximates the true posterior $P(z|x)$ with $q_\phi(z|x)$

$$\Rightarrow KL\left(q_\phi(z) \,\|\, P_\theta(z|x)\right) = \int q_\phi(z) \log\left[\frac{q_\phi(z)}{P_\theta(z|x)}\right]\frac{P(x)}{P(x)}\,dz$$

Using $q(z)$ to approximate $P(z|x)$

$$= -\int q_\phi(z) \log \frac{P(x,z)}{P(x)\,q_\phi(z)}\,dz = -\int q(z) \log \frac{P(x,z)}{q_\phi(z)}\,dz + \int q_\phi(z) \log P(x)\,dz$$

$$= \underbrace{-E_{q_\phi(z)}\left[\log \frac{P(x,z)}{q_\phi(z)}\right]}_{\text{ELBO}} + \log P(x)$$

ELBO (Method 1)

we started with this term

Real posterior (unknown)

$$\Rightarrow \log P(x) = \text{ELBO} + KL\left[q_\phi(z) \,\|\, P_\theta(z|x)\right]$$

Always positive

If we drop the KL term we bound $\log P(x)$

$$\log P(x) \geq E_{q_\phi(z)}\left[\log \frac{P_\theta(x,z)}{q_\phi(z)}\right]$$

Higher ELBO → Smaller difference to true $P(z|x)$

↓ Better latent representation

Higher ELBO → gap to log-likelihood tightens

↓ Better density model

Amortized variational posterior:

$$q_\phi(z) \Rightarrow \underbrace{q_\phi(z|x)}$$

Use a neural network (encoder network)