Variational Autoencoder

ELBO (From Method 1 and 2)

$$= E_{q_\phi(z|x)}[\log \underbrace{P_\theta(x|z)}_{\text{Decoder}}] - KL[\underbrace{q_\phi(z|x)}_{\text{Encoder}} || P(z)]$$

* Standard case: approximate posterior $q(z|x)$ is Gaussian $q(z|x) = \mathcal{N}(z; \mu_z, \sigma_z)$

→ $\boxed{\text{Encoder}}$ is a standard neural network modelling the approximate posterior $q(z|x)$

→ KL term in ELBO $KL[q_\phi(z|x) || P(z)]$ encourages the posterior to match the prior $P(z)$

→ For an input x we have a distribution over latent z with μ_z & σ_z and not just single values

→ Shared NN architecture with two outputs: μ_z, σ_z

→ $\boxed{\text{Decoder}}$: NN with input Z and returns the generated output.

→ Choose appropriate distribution for the output based on the data type $P(x|z)$

Ex: continuous values → Gaussian

For images $x \in \{0, 1, 2, ... 255\}^D$ → Categorical

→ Another interpretation for $KL[q_\phi(z|x) || P(z)]$

This can be seen as a regularizer. For more complex models this regularizer may not be interpreted as the KL term.

→ Often prior is unit Gaussian $P(z) \sim N(0,1)$

→ If we desire different nature of z; pick prior accordingly. Typically, $P(z) = \mathcal{N}(0,1)$

$$q_\phi(z|x) = \mathcal{N}(\underbrace{\mu_\phi(x)}, \text{diag}(\underbrace{\sigma^2_\phi(x)}))$$

Encoder output

$\text{len}(\mu_\phi(x)) = \text{len}(\text{diag}(\sigma^2_\phi(x)))$