Models	Observed	Latent	Joint factorization/ Graphical model
FA	$\vec{x} \in \cancel{\{0,1\}^{784}} \mathbb{R}^D$	$\vec{z} \in \mathbb{R}^L$	$p_\theta(\vec{x}, \vec{z}) = \underbrace{p_\theta(\vec{x}\mid\vec{z})}_{N} \underbrace{p_\theta(\vec{z})}_{N(0,I)}$
VAE	$\vec{x} \in \{0,1\}^{784}$	$\vec{z} \in \mathbb{R}^L$	$p_\theta(\vec{x}, \vec{z}) = \underbrace{p_\theta(\vec{x}\mid\vec{z})}_{\text{ProdBern}} \underbrace{p(\vec{z})}_{N(0,I)}$
GMVAE	$\vec{x} \in \{0,1\}^{784}$	$\vec{y}$ (one-hot), $\vec{z} \in \mathbb{R}^L$	$p_\theta(\vec{x},\vec{y},\vec{z}) = \underbrace{p_\theta(\vec{x}\mid\vec{z})}_{\text{ProdBern}} \underbrace{p_\theta(\vec{z}\mid\vec{y})}_{N} \underbrace{p(\vec{y})}_{\text{OneHot Cat}}$
VRNN	$\vec{x}_{1:T}$, each $\vec{x}_t \in \{0,1\}^{28}$	$\vec{z}_{1:T}$, each $\in \mathbb{R}^L$	$p_\theta(\vec{x}_{1:T}, \vec{z}_{1:T}) = \prod_{t=1}^{T} \overbrace{p_\theta(\vec{x}_t \mid \vec{x}_{<t}, \vec{z}_{\le t})}^{\text{ProdBern}} \underbrace{p_\theta(\vec{z}_t \mid \underbrace{\vec{x}_{<t}, \vec{z}_{<t}}_{h_t})}_{N}$

$\nabla_\phi, \nabla_\theta$

single data point: i

$$\log p_\theta(\text{observed}_i) \ge \mathbb{E}_{q_\phi(\text{latent}/\text{observed})}\left[\log p_\theta(\text{obs}, \text{latent}) - \log q_\phi(\text{latent} \mid \text{obs})\right]$$

$p_\theta(\text{latent}/\text{observed})$

Models	True Post / Its factorization	Approx post
FA	$p_\theta(\vec{z}\mid\vec{x}) = N(Vx, \Sigma_{II})$ can be obtained using complicated formula	$q_\phi(\vec{z}\mid\vec{x})$ takes the form of $N(Vx, \Sigma)$ where $\phi = (V, \Sigma)$
VAE	$p_\theta(\vec{z}\mid\vec{x})$ but not sure about its form	$q_\phi(\vec{z}\mid\vec{x}) = N(\vec{\mu}_\phi(\vec{x}), \vec{\sigma}_\phi(\vec{x}))$
GMVAE	$p_\theta(\vec{y},\vec{z}\mid\vec{x}) = p_\theta(\vec{y}\mid\vec{x})\, p_\theta(\vec{z}\mid\vec{x},\vec{y})$ onehot-categorical; not known	$q_\phi(\vec{y}\mid\vec{x}) = \text{OneHotCat}(\pi_\phi(\vec{x}))$ $q_\phi(\vec{z}\mid\vec{x},\vec{y}) = N(\vec{\mu}_\phi(\vec{x},\vec{y}), \vec{\sigma}_\phi(\vec{x},\vec{y}))$
VRNN		

$$p_\theta(z_{1:T} \mid x_{1:T}) = p_\theta(z_1 \mid x_{1:T})\, p_\theta(z_2 \mid z_1, x_{1:T}) \cdots$$

$$q_\phi(z_1 \mid x_1)\, q_\phi(z_2 \mid z_1, x_{1:2}) \cdots$$

$$p_\theta(z_2 \mid x_1, z_1)$$

Derivation of unbiased estimator of gradient for GMVAE

$$\nabla_\phi \mathbb{E}_{q_\phi(\vec{y},\vec{z}|\vec{x})}\left[\log p_\theta(\vec{x},\vec{y},\vec{z}) - \log q_\phi(\vec{y},\vec{z}|\vec{x})\right]$$

$$= \nabla_\phi \mathbb{E}_{q_\phi(\vec{y}|\vec{x})\, q_\phi(\vec{z}|\vec{y},\vec{x})}\Big[\log p_\theta(\vec{x}|\vec{z}) + \log p_\theta(\vec{z}|\vec{y}) + \log p(\vec{y}) - \log q_\phi(\vec{y}|\vec{x}) - \log q_\phi(\vec{z}|\vec{x},\vec{y})\Big]$$

$$= \left(\nabla_\phi \cancel{\mathbb{E}_{q_\phi(\vec{y})}} \sum_{\vec{y}} q_\phi(\vec{y}|\vec{x}) \left[\mathbb{E}_{q_\phi(\vec{z}|\vec{y},\vec{x})}[\cdots]\right]\right) + \nabla_\phi \mathbb{E}_{q_\phi(\vec{y}|\vec{x})} \underbrace{\left[\log p(\vec{y}) - \log q_\phi(\vec{y}|\vec{x})\right]}_{\text{closed-form}}$$

$$= \left(\sum_{\vec{y}} \nabla_\phi\{q_\phi(\vec{y}|\vec{x})\} \cdot \mathbb{E}_{q_\phi(\vec{z}|\vec{y},\vec{x})}[\cdots] + q_\phi(\vec{y}|\vec{x})\, \nabla_\phi \mathbb{E}_{q_\phi(\vec{z}|\vec{y},\vec{x})}[\cdots]\right) + \cdots$$

evaluate this with $\vec{z}^s \sim q_\phi(\vec{z}|\vec{y},\vec{x})$

evaluate this with $\vec{z}^s \sim q_\phi(\vec{z}|\vec{y},\vec{x})$ but sample it using reparametrization trick

$\simeq$

$\cancel{\sum_{\vec{y}} \nabla_\phi\{q_\phi(\vec{y}|\vec{x})\}}$

Der.

$$\nabla_\phi \mathbb{E}_{q_\phi(\vec{z}|\vec{x})}\left[\log p_\theta(\vec{x},\vec{z}) - \log q_\phi(\vec{z}|\vec{x})\right]$$

$$= \nabla_\phi\left\{\mathbb{E}_{q_\phi(\vec{z}|\vec{x})}\left[\log p_\theta(\vec{x}|\vec{z}) + \log p_\theta(\vec{z}) - \log q_\phi(\vec{z}|\vec{x})\right]\right\}$$

$$= \nabla_\phi \mathbb{E}_{q_\phi(\vec{z}|\vec{x})}\left[\log p_\theta(\vec{x}|\vec{z})\right] + \nabla_\phi \mathbb{E}_{q_\phi(\vec{z}|\vec{x})}\underbrace{\left[\log p(\vec{z}) - \log q_\phi(\vec{z}|\vec{x})\right]}_{-\,\mathrm{KL}\ \text{closed form}}$$

evaluate this with $\vec{z}^s \sim q_\phi(\vec{z}|\vec{x})$ with reparametrization

$\Rightarrow$ (1) evaluate $\log p_\theta(\vec{x}|\vec{z}^s) + D_{KL}(p(\vec{z}) - \log q_\phi(\vec{z}|\vec{x}))$ (2) compute gradient