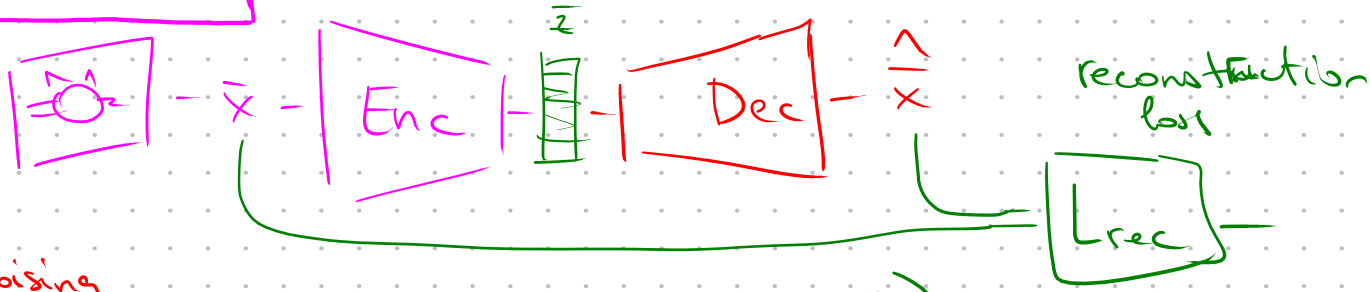
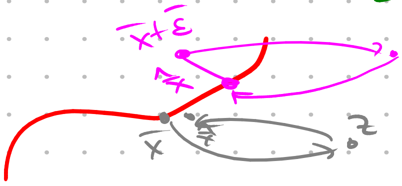
Autoencoders



denoising autoencoder

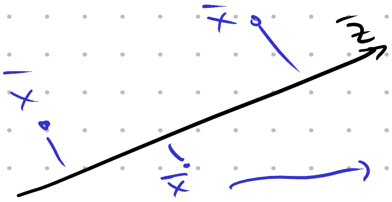
$$L_{rec} = d(\bar{x}, Dec(Enc(\bar{x})))$$

$$L_{rec} = d(\bar{x}, Dec(Enc(\bar{x}+\bar{\varepsilon})))$$



PCA - principal components analysis

ICA independent components analysis

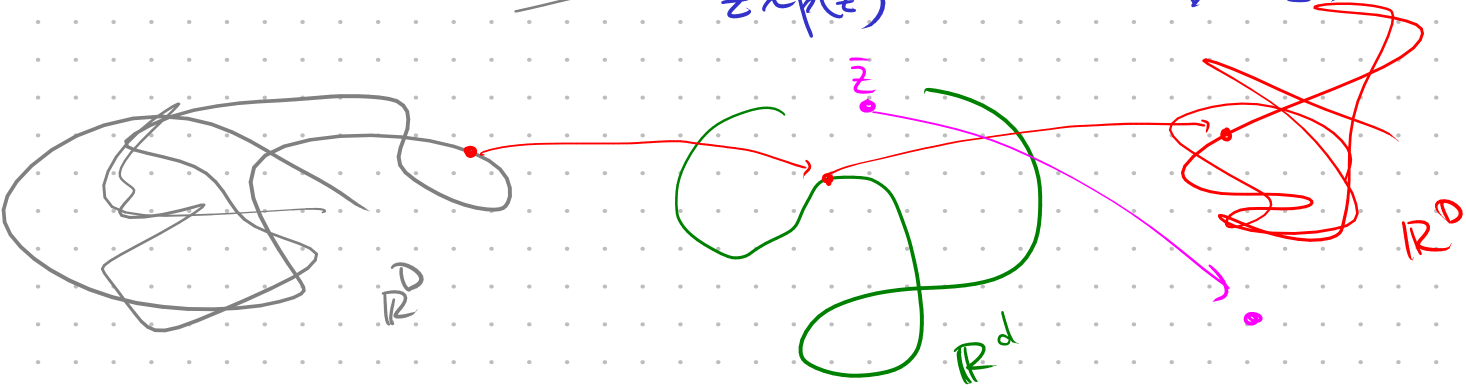


$$\underset{Enc(\bar{x})}{\bar{z} = W\bar{x}} \quad \rightarrow \quad \underset{Dec(\bar{x})}{\hat{x} = W^T\bar{z}}$$

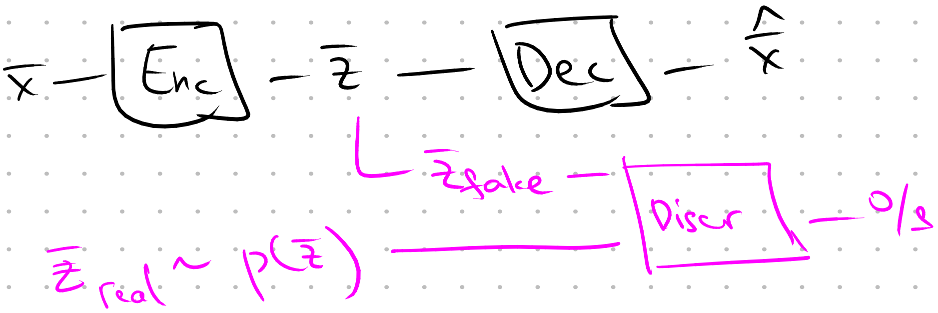
$$\boxed{\sum_n \|W^T W\bar{x} - \bar{x}\|_2^2 \xrightarrow[W]{} \min}$$

$\bar{x} \in \mathbb{R}^D$ — Enc — $\bar{z} \in \mathbb{R}^d$ — Dec — $\hat{x} \in \mathbb{R}^D$

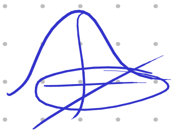
$\bar{z} \sim p(\bar{z})$ → $Dec(\bar{z}) \in \mathbb{R}^D$



AAE

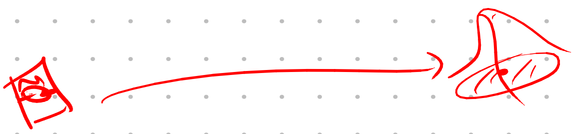


VAE – variational autoencoder



$$\overset{\mathcal{N}(\bar z|\bar 0, I)}{\overset{\|}{p(\bar z)}} \sim \bar z \longrightarrow \boxed{Dec} \longrightarrow Dec(\bar z)$$

$$\bar x - \boxed{Enc} - \underline{p(\bar z|\bar x)} \sim \bar z \longrightarrow \boxed{Dec} \longrightarrow \hat{\bar x}$$



$\|\bar x - \hat{\bar x}\|_2^2$

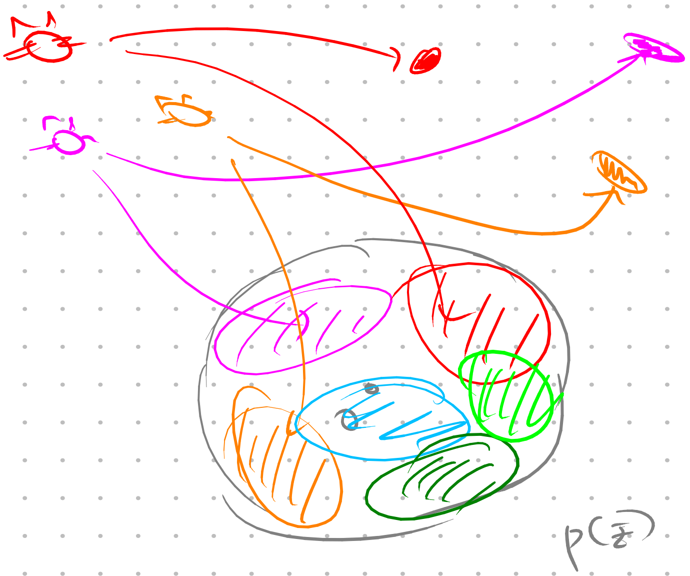
$\boxed{L_{rec}}$

$$\bar x - \boxed{Enc} - \bar\theta_x = (\bar\mu_x, \bar\sigma_x)$$

$$\bar z \sim \underline{p(\bar z|\bar\theta_x)} = \mathcal{N}(\bar z|\bar\mu_x, \bar\sigma_x) \longrightarrow \boxed{Dec} \longrightarrow \hat{\bar x}$$

$\mathcal{N}(\bar 0, I)$

$\boxed{L_{reg}}$

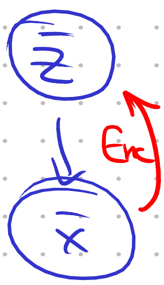


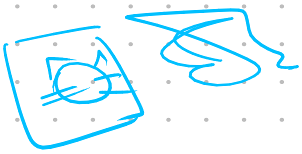
$$\boxed{L = L_{rec} + \alpha L_{reg}}$$

$$L(\bar x) = \|\bar x - \hat{\bar x}\|_2^2 + \alpha \cdot KL\big(\mathcal{N}(\bar z|\bar\mu_x, \bar\sigma_x)\,\|\,\mathcal{N}(\bar z|\bar 0, I)\big)$$

VAE

$$p(\bar x)\,\underbrace{p(\bar z|\bar x)}_{Enc(\bar x)} = p(\bar x, \bar z) = \underbrace{p(\bar z)}_{\mathcal{N}(\bar z|\bar 0, I)}\,\underbrace{p(\bar x|\bar z)}_{Dec(\bar z)}$$





$$p(\bar x|\bar z) = \mathcal{N}(\bar x|Dec(\bar z), cI)$$

$$p(\bar z|\bar x) = \frac{p(\bar z)\,p(\bar x|\bar z)}{p(\bar x)} \approx q(\bar z) = \mathcal{N}(\bar z|\bar\mu_x, \bar\sigma_x)$$

Variational approx.

$$p(\bar{x},\bar{z}) = p(\bar{x})\, p(\bar{z}|\bar{x})$$

$$E_{q(\bar{z})}\left[\log p(\bar{x})\right] = E_{q(\bar{z})}\left[\log p(\bar{x},\bar{z}) - \log p(\bar{z}|\bar{x})\right]$$

$$\log p(\bar{x}) = \int q(\bar{z}) \log p(\bar{x},\bar{z})\, d\bar{z} - \int q(\bar{z}) \log p(\bar{z}|\bar{x})\, d\bar{z}$$

$$\pm \int q \log q \, dz$$

$$\underline{\log p(\bar{x})} = \boxed{\int q(\bar{z}) \log \frac{p(\bar{x},\bar{z})}{q(\bar{z})} d\bar{z}} - \underline{\int q(\bar{z}) \log \frac{p(\bar{z}|\bar{x})}{q(\bar{z})} d\bar{z}}$$

$$\text{const} = \mathcal{L}(q) + \underline{KL\left(q(\bar{z}) \,\|\, p(\bar{z}|\bar{x})\right)}$$

$$\xrightarrow[q]{} \max \qquad \xrightarrow[q]{} \min$$

$$p(\bar{z})\, p(\bar{x}|\bar{z})$$

$$\mathcal{L}(q) = \int q(\bar{z}) \log \frac{p(\bar{x},\bar{z})}{q(\bar{z})} d\bar{z} =$$

$$= \underline{\int q(\bar{z}) \log p(\bar{x}|\bar{z})\, d\bar{z}} + \boxed{\int q(\bar{z}) \log \frac{p(\bar{z})}{q(\bar{z})} d\bar{z}}$$

$$= \underline{\int q(\bar{z}) \log p(\bar{x}|\bar{z})\, d\bar{z}} - KL\left(q(\bar{z}) \,\|\, p(\bar{z})\right) \xrightarrow[q]{} \max$$

$$\mathcal{N}(\bar{x} \,|\, f(\bar{z}), cI)$$

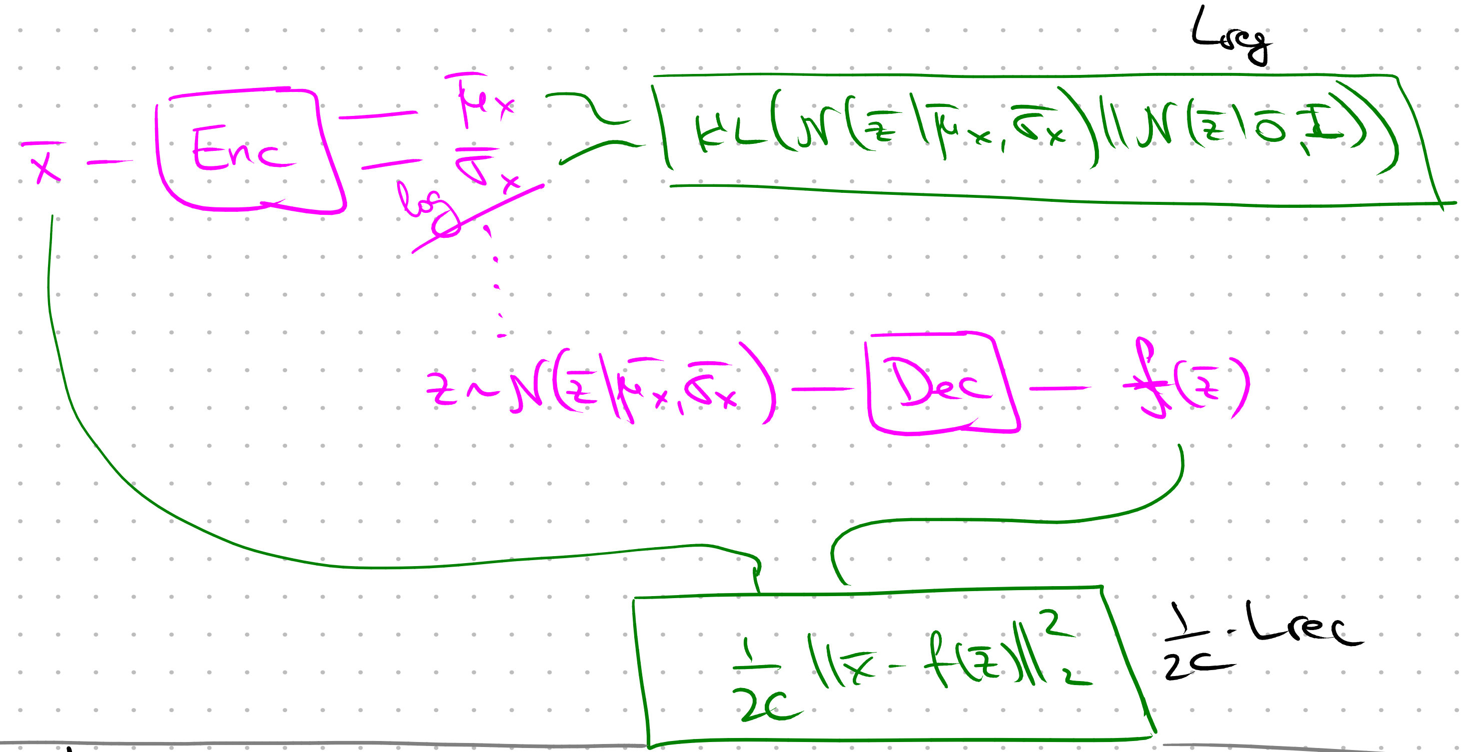
$$\log p(\bar{x}|\bar{z}) = \text{const} - \underline{\frac{1}{2c}(\bar{x} - f(\bar{z}))^T(\bar{x} - f(\bar{z}))}$$

$$\mathcal{N}(\bar{z} \,|\, \bar{\mu}_x, \bar{\sigma}_x)$$

$$q \to \min$$

$$+ \underline{E_{q(\bar{z})}\left[\frac{1}{2c}\|\bar{x} - f(\bar{z})\|^2\right]} + \underline{KL\left(q(\bar{z}) \,\|\, p(\bar{z})\right)} \xrightarrow[q]{} \max$$

$$\frac{1}{2c} \cdot L_{rec} + L_{reg}$$



L_{reg}

$$KL\left(\mathcal{N}(\bar z|\bar\mu_x,\bar\sigma_x)\,\|\,\mathcal{N}(\bar z|\bar 0, I)\right) =$$

$$\bar\sigma_x = \begin{pmatrix}\sigma_x^1 & & 0\\ & \ddots & \\ 0 & & \sigma_x^d\end{pmatrix}$$

$$= KL\Big(\prod_{i=1}^{d} \underbrace{q_i(z_i)}_{=\mathcal{N}(z_i|\mu_{xi},\sigma_{xi})} \,\Big\|\, \prod_{i=1}^{d} \underbrace{p_i(z_i)}_{=\mathcal{N}(z_i|0,1)}\Big) =$$

$$= \int \prod_i q_i \cdot \log\frac{\prod q_i}{\prod p_i}\, d\bar z = \int \prod_{i=1}^{d} q_i \left(\log\frac{q_1}{p_1} + \log\frac{q_2}{p_2} + \dots + \log\frac{q_d}{p_d}\right) d\bar z =$$

$$= \sum_{j=1}^{d} \int \log\frac{q_j(z_j)}{p_j(z_j)} \cdot \prod_{i=1}^{d} q_i(z_i)\, dz_1\, dz_2 \dots dz_d =$$

$$= \sum_{j=1}^{d} \left(\left(\int \log\frac{q_j(z_j)}{p_j(z_j)}\, q_j(z_j)\, dz_j\right) \cdot \underbrace{\prod_{i\neq j} \int q_i(z_i)\, dz_i}_{=1}\right) =$$

$$= \sum_{j=1}^{d} KL\left(q_j(z_j)\,\|\,p_j(z_j)\right) =$$

$$= \sum_{j=1}^{d} KL\left(\mathcal{N}(z_j \mid \mu_{x_j}, \sigma_{x_j}) \,\|\, \mathcal{N}(z_j \mid 0, 1) \right) =$$

$$= \sum_{j=1}^{d} \underbrace{\int \frac{1}{\sqrt{2\pi\sigma_{x_j}^2}} e^{-\frac{1}{2\sigma_{x_j}^2}(z_j - \mu_{x_j})^2}}\, \cdot \log \frac{\mathcal{N}(\cdot \mid \mu_{x_j}, \sigma_{x_j})}{\mathcal{N}(\cdot \mid 0, 1)} \, \underline{dz_j}$$

$$= \sum_{j=1}^{d} \mathbb{E}_{\mathcal{N}(z_j \mid \mu_{x_j}, \sigma_{x_j})} \Big[-\cancel{\tfrac{1}{2}\log 2\pi} - \tfrac{1}{2}\log \sigma_{x_j}^2 - \frac{1}{2\sigma_{x_j}^2}(z_j - \mu_{x_j})^2$$

$$+ \cancel{\tfrac{1}{2}\log 2\pi} + \tfrac{1}{2} z_j^2 \Big] =$$

$$= \sum_{j=1}^{d} \mathbb{E}_{\mathcal{N}(z_j \mid \mu_{x_j}, \sigma_{x_j})} \left[-\frac{1}{2}\log \sigma_{x_j}^2 - \frac{\mu_{x_j}^2}{2\sigma_{x_j}^2} + \left(\frac{1}{2} - \frac{1}{2\sigma_{x_j}^2}\right) z_j^2 + \frac{\mu_{x_j}}{\sigma_{x_j}^2} z_j \right]$$

$$\mathbb{E}[z_j] = \mu_{x_j}$$

$$\mathbb{E}[z_j^2] = (\mathbb{E}[z_j])^2 + \mathrm{Var}\, z_j = \mu_{x_j}^2 + \sigma_{x_j}^2$$

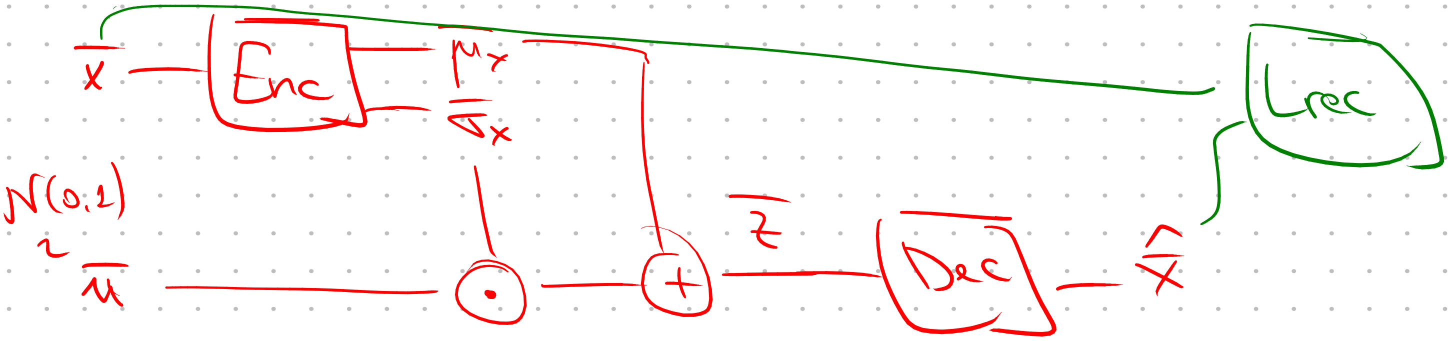
$$= \sum_{j=1}^{d} \left(-\frac{1}{2}\log \sigma_{x_j}^2 - \cancel{\frac{\mu_{x_j}^2}{2\sigma_{x_j}^2}} + \underbrace{\frac{1}{2}\left(1 - \frac{1}{\sigma_{x_j}^2}\right)(\mu_{x_j}^2 + \sigma_{x_j}^2)} + \cancel{\frac{\mu_{x_j}}{\sigma_{x_j}^2}} \cdot \mu_{x_j} \right)$$

$$+ \frac{1}{2}\mu_{x_j}^2 + \frac{1}{2}\sigma_{x_j}^2 - \cancel{\frac{\mu_{x_j}^2}{2\sigma_{x_j}^2}} - \frac{1}{2}$$

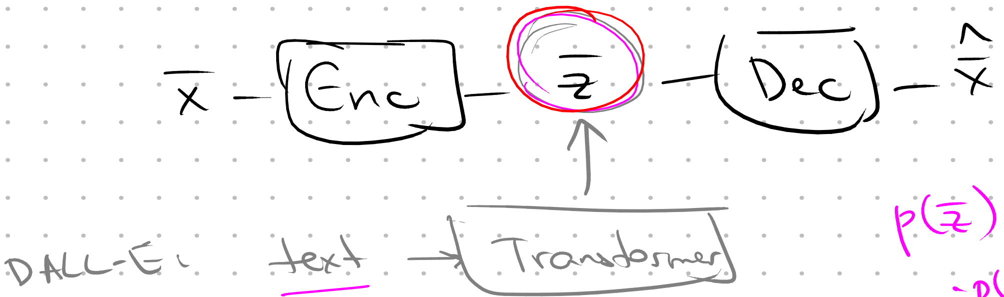
$$= \sum_{j=1}^{d} \frac{1}{2}\left(\sigma_{x_j}^2 + \mu_{x_j}^2 - \log \sigma_{x_j}^2 - 1 \right)$$

Reparametrization trick

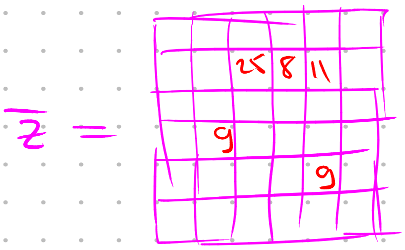
$$z_j \sim N(z_j \mid \mu_{x_j}, \sigma_{x_j}) \iff \begin{array}{l} u_j \sim N(0,1) \\ z_j = \mu_{x_j} + \sigma_{x_j} \cdot u_j \end{array}$$



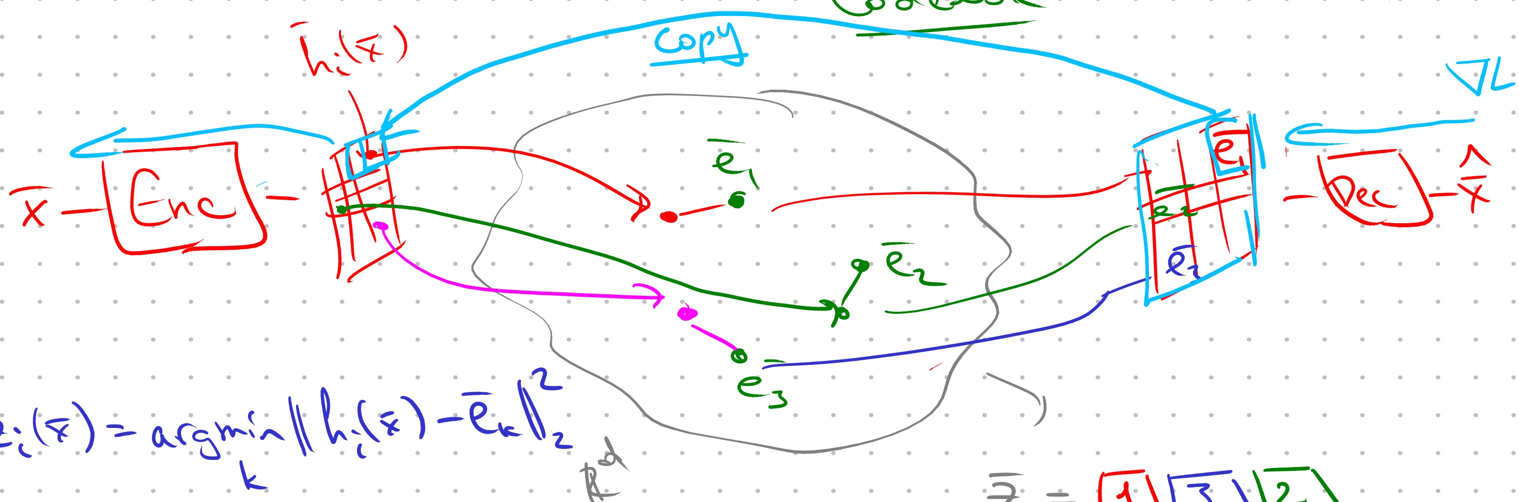
VQ-VAE – Vector quantized VAE



$$p(\bar{z}) = p(z_1)\,p(z_2 \mid z_1) \cdot p(z_3 \mid z_1, z_2) \ldots p(z_d \mid z_1 \ldots z_{d-1})$$



Codebook $(\bar{e}_1, \ldots, \bar{e}_M)$



$$z_i(\bar{x}) = \arg\min_k \|h_i(\bar{x}) - \bar{e}_k\|_2^2$$

$\bar{z} =$ 1 3 2

stopgradient

$$L_{VQ\text{-}VAE} = L_{VAE} +$$

$$+ \| sg[h_e(\bar{x})] - \bar{e} \|_2^2$$

codebook alignment loss

$$+ \beta \| h_e(\bar{x}) - sg[\bar{e}] \|_2^2$$

codebook commitment loss