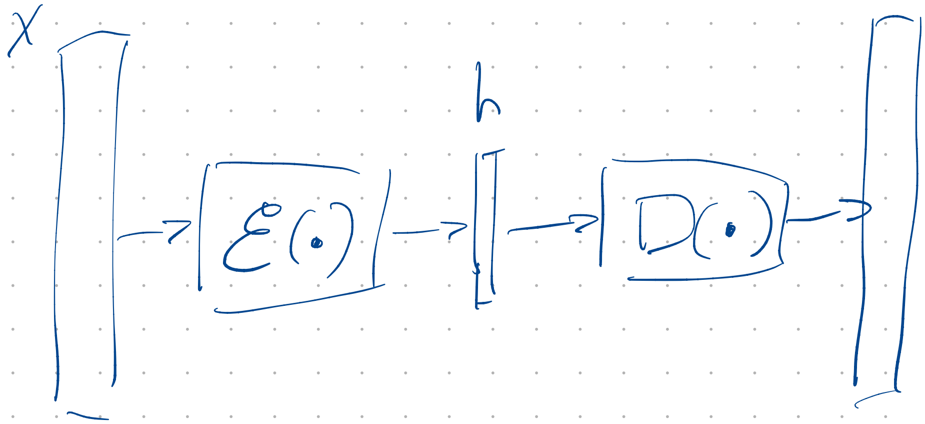
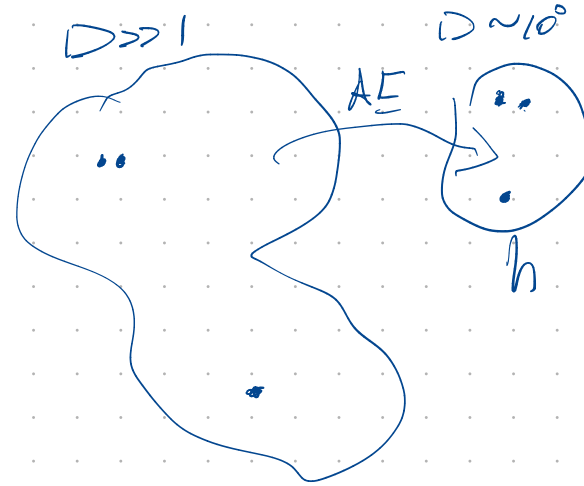


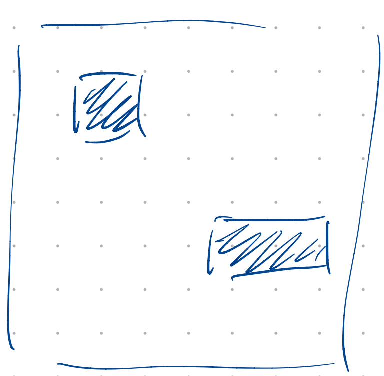


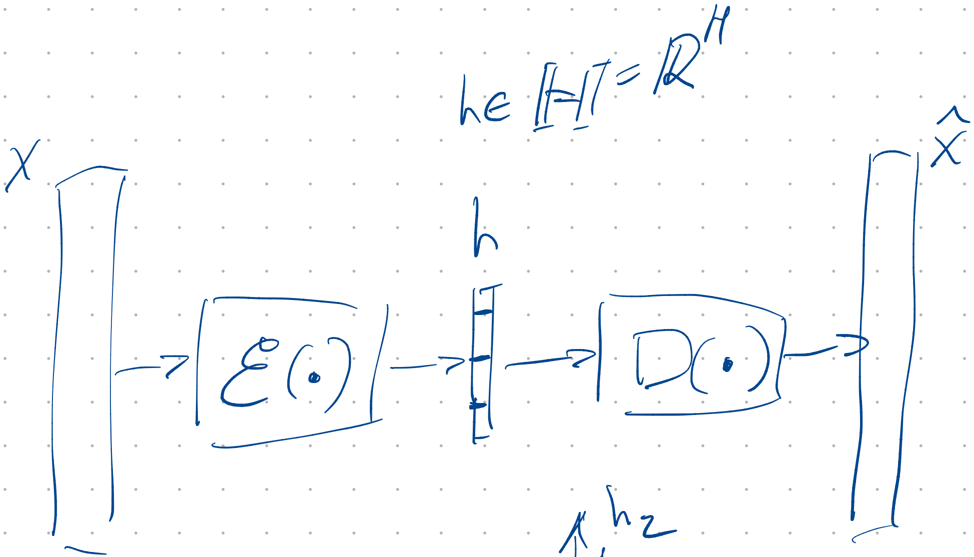
$$\hat{x} = D(\mathcal{E}(x))$$

$$\mathcal{L} = MSE(x, \hat{x})$$
$$MAE(x, \hat{x})$$
$$MAPE(x, \hat{x})$$
$$BCE(x, \hat{x})$$

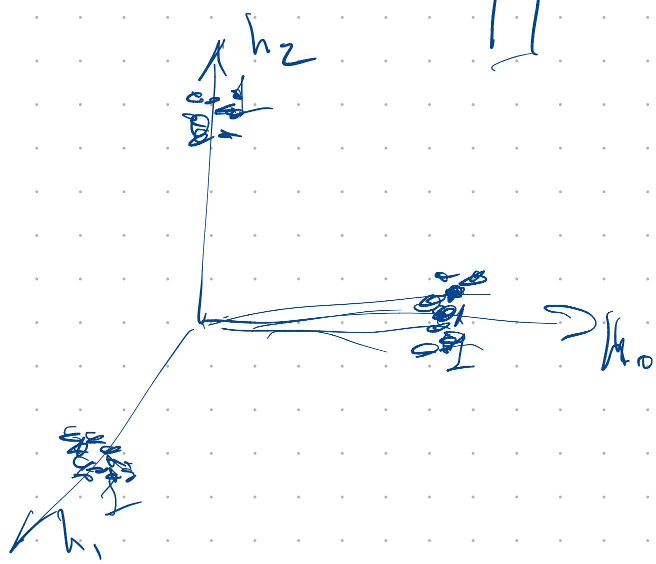
Denoising:

$$\mathcal{L} = MSE(x, D(\mathcal{E}(x + \varepsilon)))$$
$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$





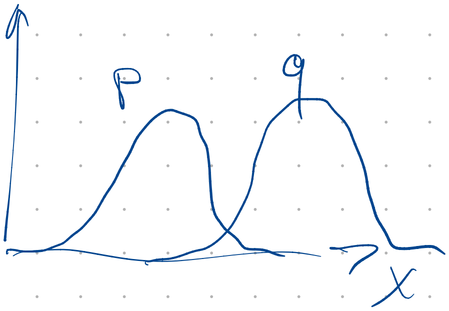
$h \in \mathbb{R}^{128}$



$$\mathcal{L} = MSE(x, \hat{x}) + \lambda \cdot reg$$

$$reg = \sum_{i=1}^{N} \sum_{j:\, h_j \geq c} (1 - h_j)^2 + \sum_{i=1}^{N} \sum_{j:\, h_j < c} h_j^2$$

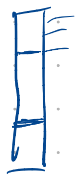
p, q $$KL(p \| q) = \int_{x \sim X} p(x) \log \frac{p(x)}{q(x)} \, dx$$



$$KL(p \| q) \neq KL(q \| p)$$

$$KL(p \| q) = -\sum p \log \frac{q}{p}$$

$h \in \mathbb{R}^H$



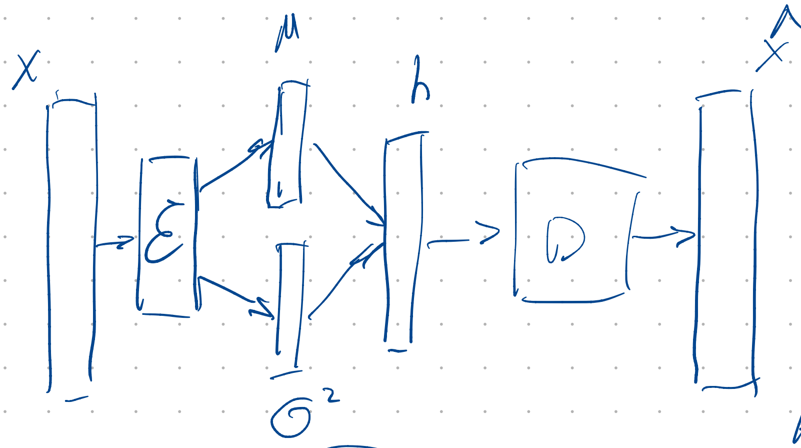
$$KL(p \| h) = -\sum p \log \frac{h}{p}$$

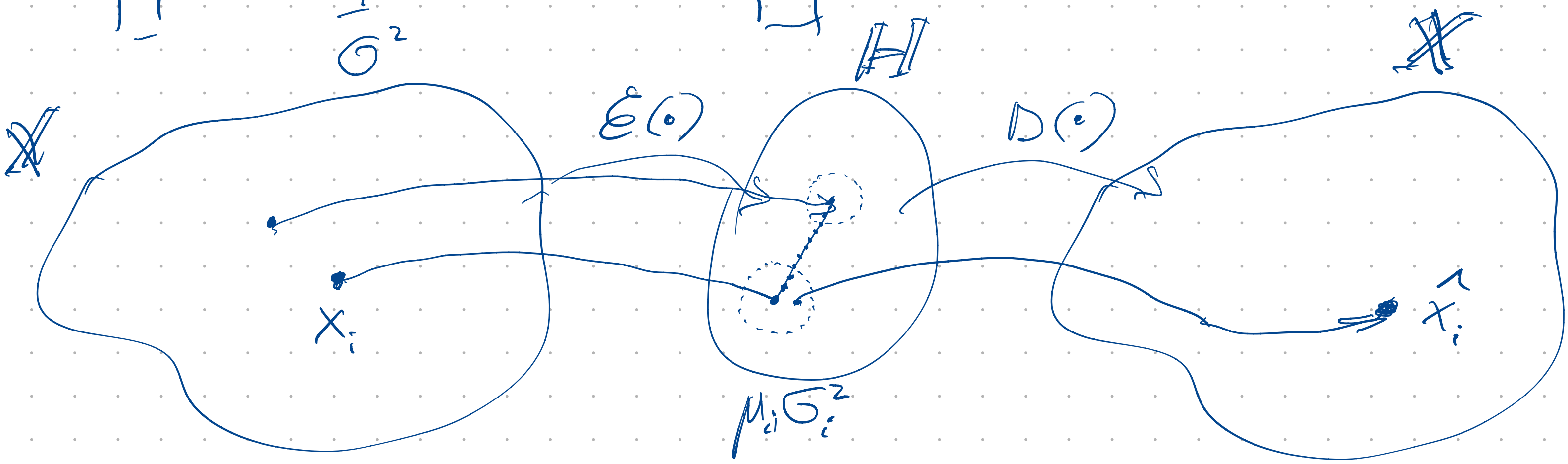
$$p = Ber(\rho) \qquad \rho = 0{,}1$$

$$KL(Ber_H(\rho), h) = -\overset{?}{} \sum_{i=1}^{N} \sum_{j=1}^{H} \left(\rho \log \frac{\rho}{h_j} + (1-\rho) \log \frac{1-\rho}{1-h_j} \right)$$

$$h = \mathcal{E}(X) \qquad reg = KL(Ber_H(\rho), h)$$

$$\mathcal{L} = MSE(x, \hat{x}) + \lambda \, KL(Ber_H(\rho), \mathcal{E}(x))$$





$$\mathcal{L} = MSE(x, \hat{x}) + \lambda \cdot reg_{VAE}$$

$$\mu_i, \sigma_i^2 = \mathcal{E}(x_i)$$

$$h_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

$$\mu, \sigma^2 \in \mathbb{R}^H$$

$$\ln(\sigma^2) \to e^{\ln \sigma^2}$$

$$\xi \sim \mathcal{N}(0, 1)$$

$$h_i = \xi \cdot \sigma_i + \mu_i$$

$$\hat{x}_i = D(h_i)$$

$$reg_{VAE} = KL(h, \mathcal{N}(\mu, \sigma^2)) =$$

$$= \frac{1}{2}\left[\sum_{i=1}^{H} \mu_i^2 + \sum_{i=1}^{H} \sigma_i^2 - \sum_{i=1}^{H} (\log \sigma_i^2 + 1)\right]$$