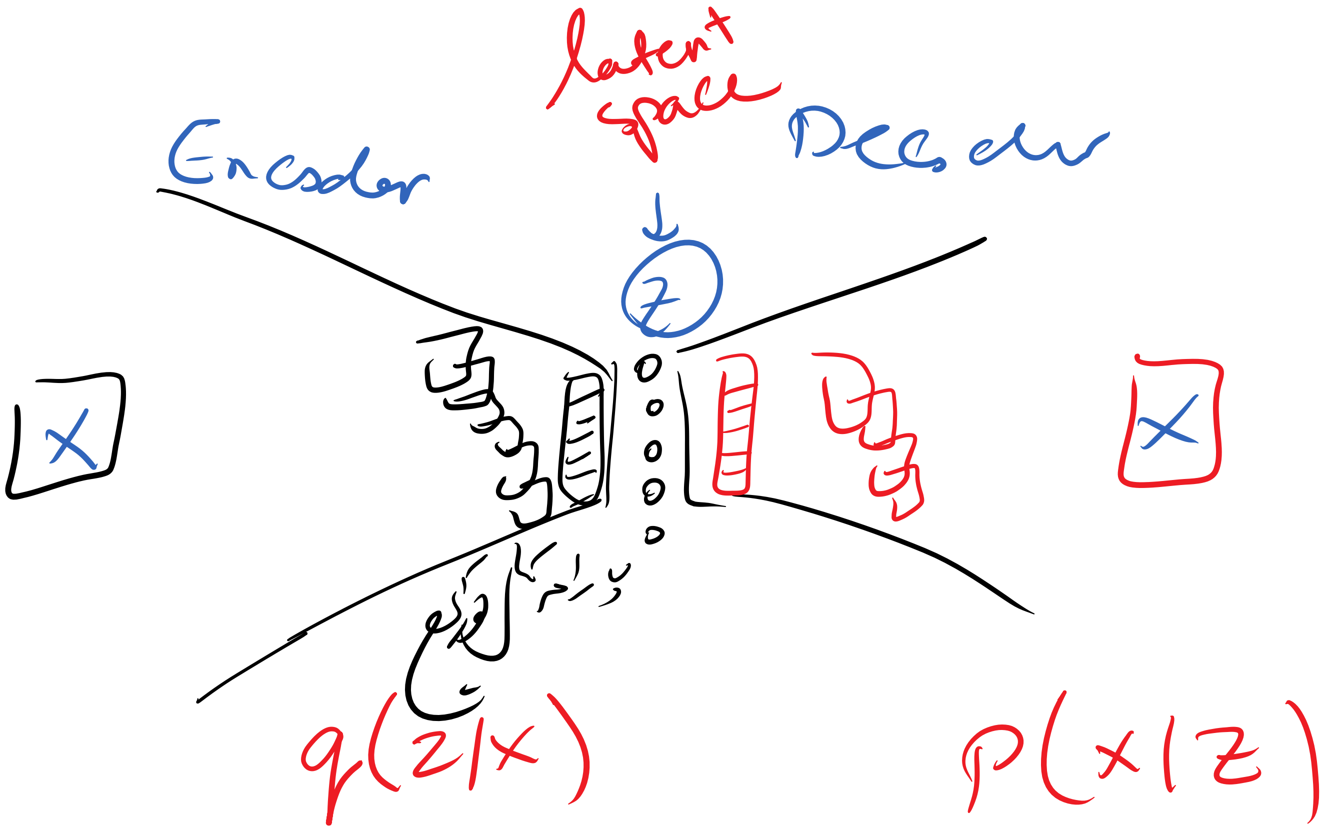
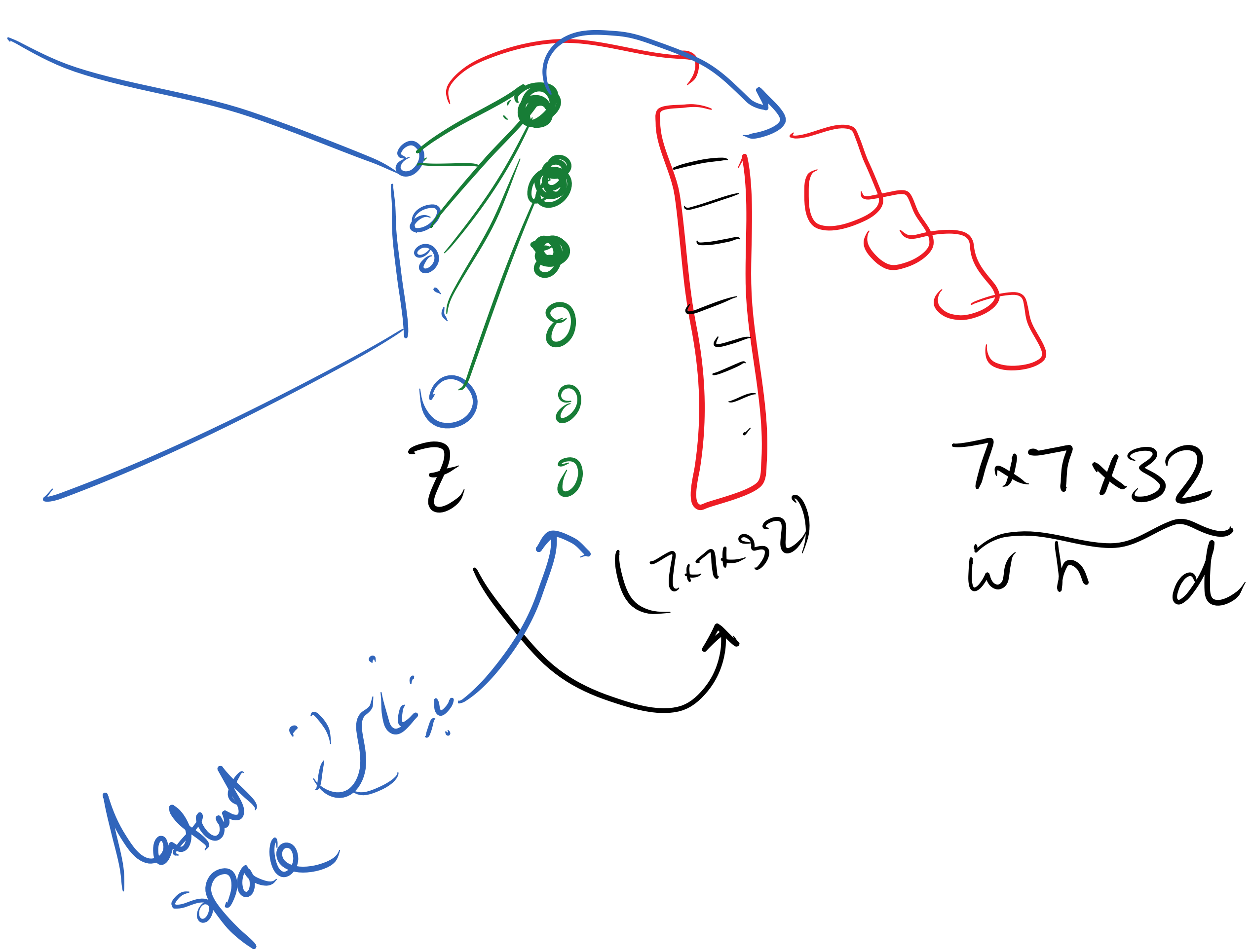
KL

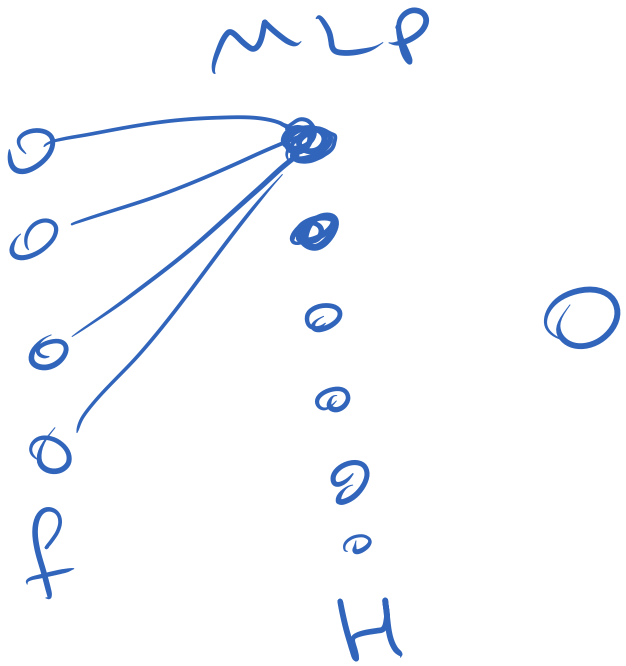
Variational AutoEncoder



$$\Sigma = \begin{bmatrix} \delta_1 & & & \\ & \delta_2 & & \phi \\ & \phi & \delta_3 & \\ & & & \ddots & \\ & & & & \delta_f \end{bmatrix}_{f \times f}$$

$$\mu = \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_f \end{bmatrix}}$$


z
(7×7×32)
7×7×32
w h d
latent space

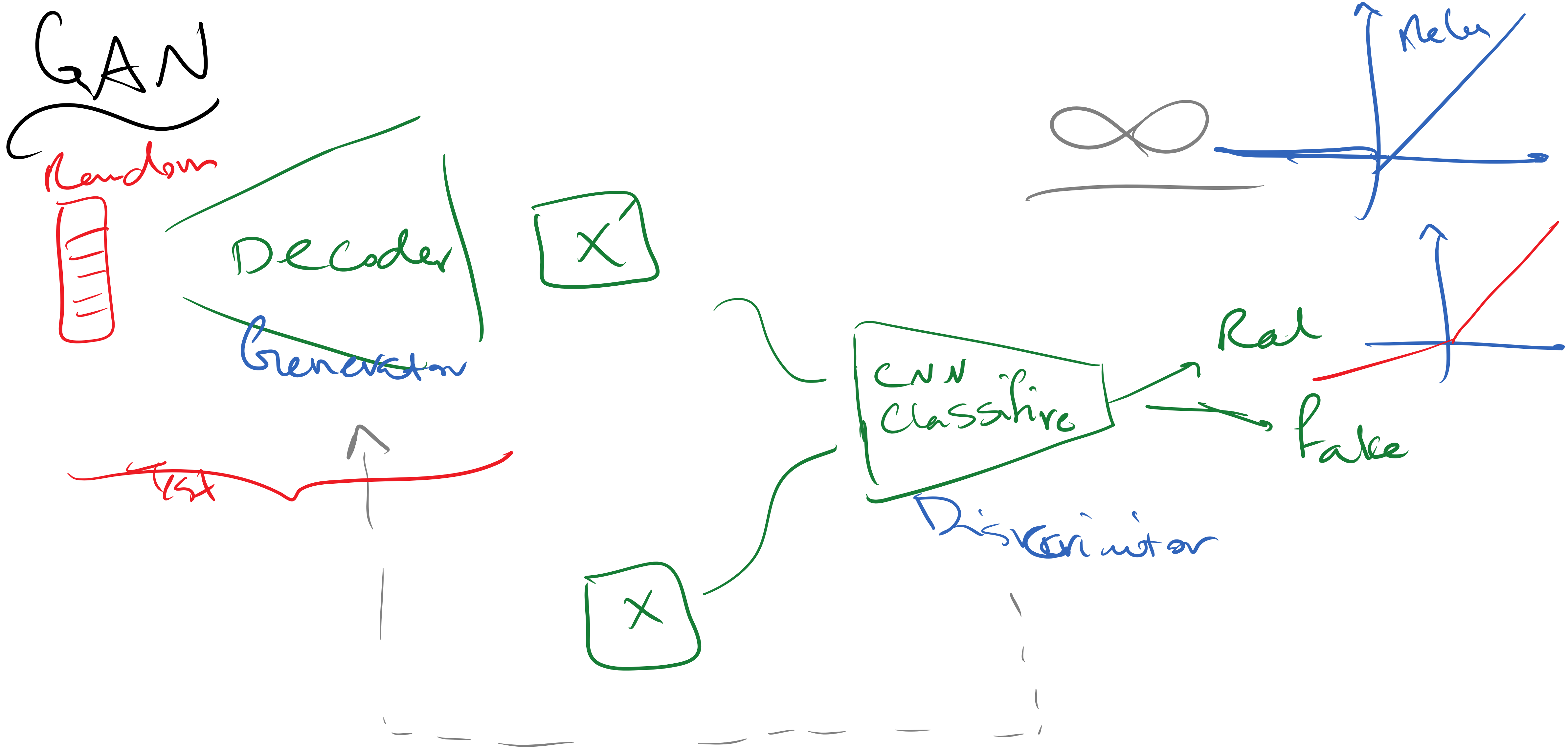

MLP
f
H

Standard Scaler

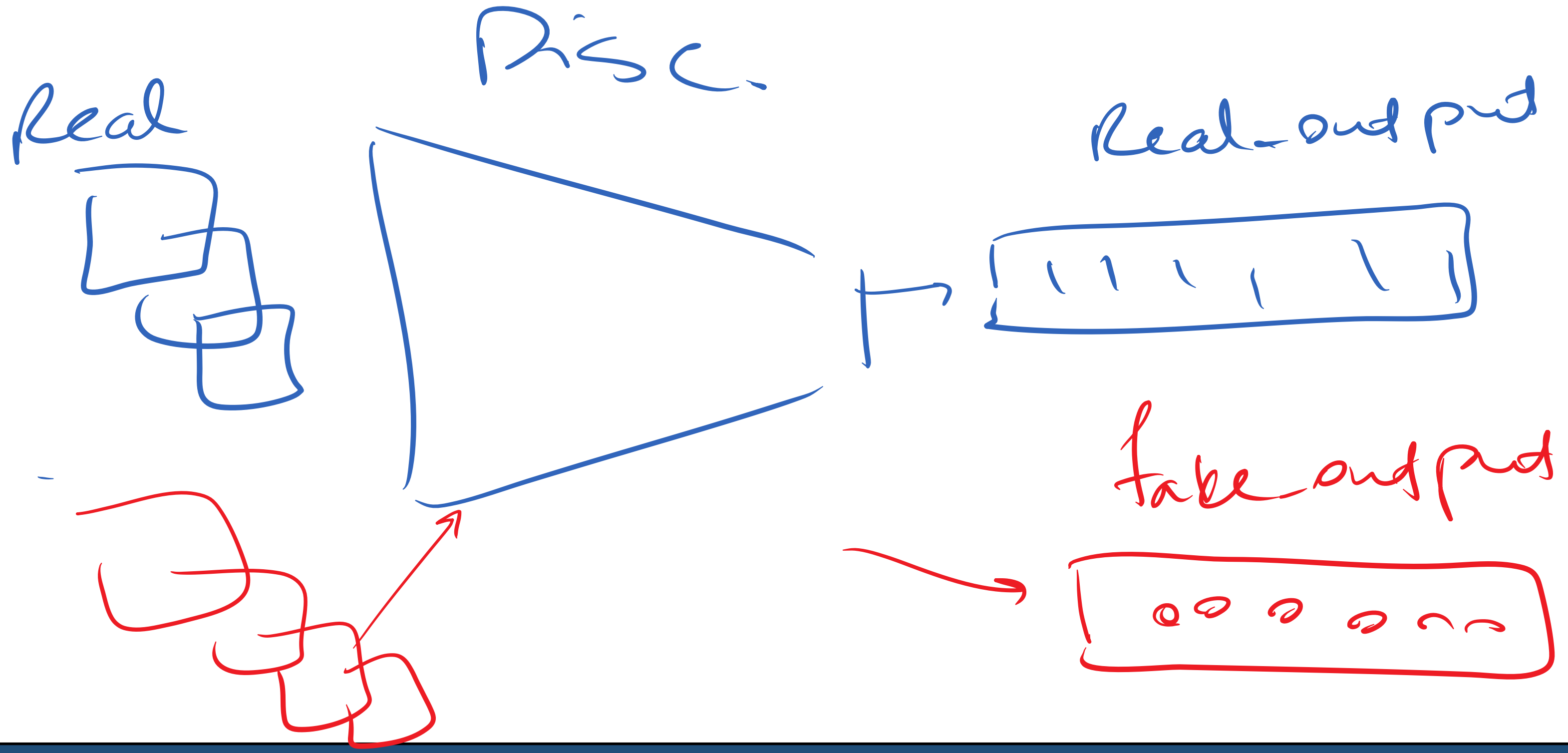
$$\mathcal{N}(\mu, \delta) \xrightarrow{\frac{x-\mu}{\delta}} \mathcal{N}(0,1)$$

$$\mathcal{N}(\mu, \delta) \quad z = \underline{\mu} + \underline{\delta} \times \varepsilon$$

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

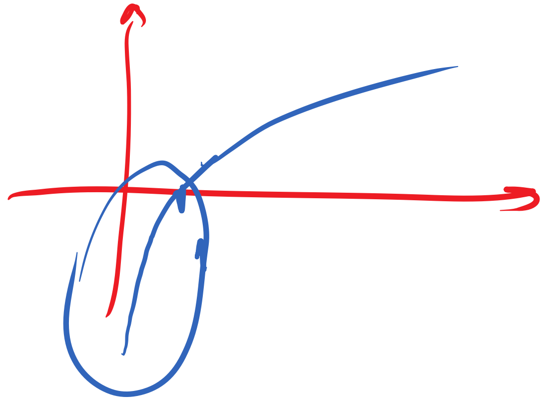

GAN
Random
Decoder
Generator
X
X
CNN
Classifire
Real
Fake
Discriminator
Relu

$$w_N = w_o - \eta \frac{\partial L}{\partial w}$$



Information : $-\log p(x)$

تعجب



Entropy : $-\sum p(x) \log p(x)$

$$KL_{(p,q)} : -\sum \overset{p(x)}{\cancel{q(x)}} \log(q(x)) + \sum p(x) \log p(x)$$

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

$KL \geq 0$

$$z \rightarrow x$$

$$\underbrace{P(z|x)}_{\text{توزیع متغیرهای نهان به شرط داده ها}} = \frac{P(x,z)}{P(x)} \longrightarrow \underbrace{q(z)}_{\text{توزیع [illegible] متغیرهای نهان}}$$

$$\text{Min } KL\big(q(z) \,||\, p(z|x)\big) = -\sum q(z) \log \frac{p(z|x)}{q(z)}$$

$$= -\sum q(z) \log \frac{\frac{p(x,z)}{p(x)}}{q(z)} = -\sum q(z) \log\left(\frac{p(x,z)}{q(z)} \times \frac{1}{p(x)}\right)$$

$$-\sum q(z)\log\left(\frac{P(x,z)}{q(z)}\times\frac{1}{P(x)}\right) = -\sum q(z)\left[\underbrace{\log\frac{P(x,z)}{q(z)}}-\underbrace{\log P(x)}\right]$$

$$= -\sum_z q(z)\log\frac{(x,z)}{q(z)} + \underbrace{\sum_z q(z)\log(P(x))}$$

$$-\sum q(z)\log\frac{(x,z)}{q(z)} + \log P(x)\underbrace{\sum_z q(z)}_{1}$$

$$KL(q(z)\,\|\,P(z|x)) = -\sum q(z)\log\frac{(x,z)}{q(z)} + \log P(x)$$

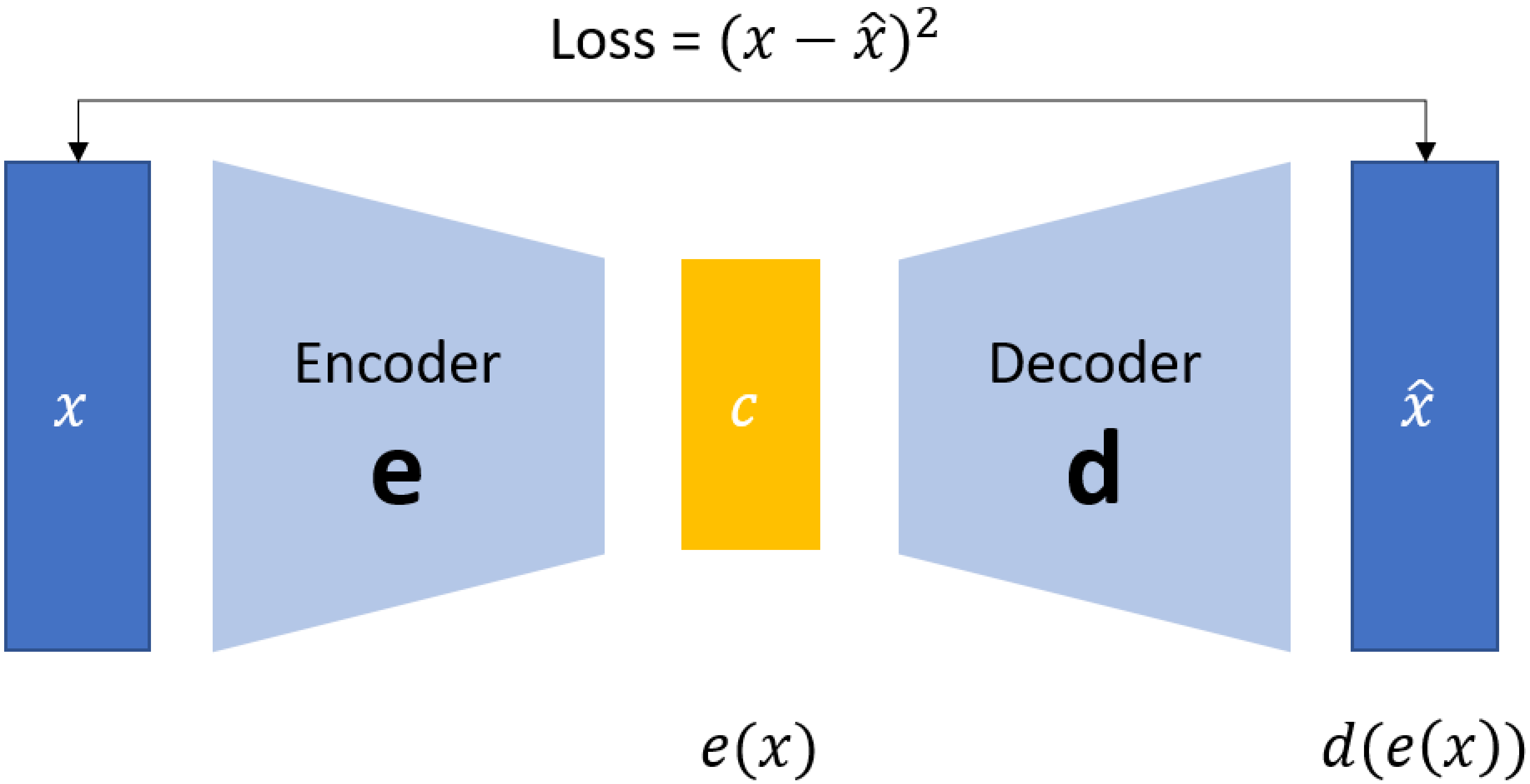
$$KL\big(q(z)\,||\,p(z|x)\big) = -\sum q(z)\log\frac{(x,z)}{q(z)} + \log p(x)$$

$$\underbrace{KL\big(q(z)\,||\,p(z|x)\big)}_{\downarrow} + \underbrace{\sum_z q(z)\log\frac{(x,z)}{q(z)}}_{\substack{\text{ELBO}\\ \mathcal{L}}} = \underbrace{\log(p(x))}_{\text{fixed}}$$

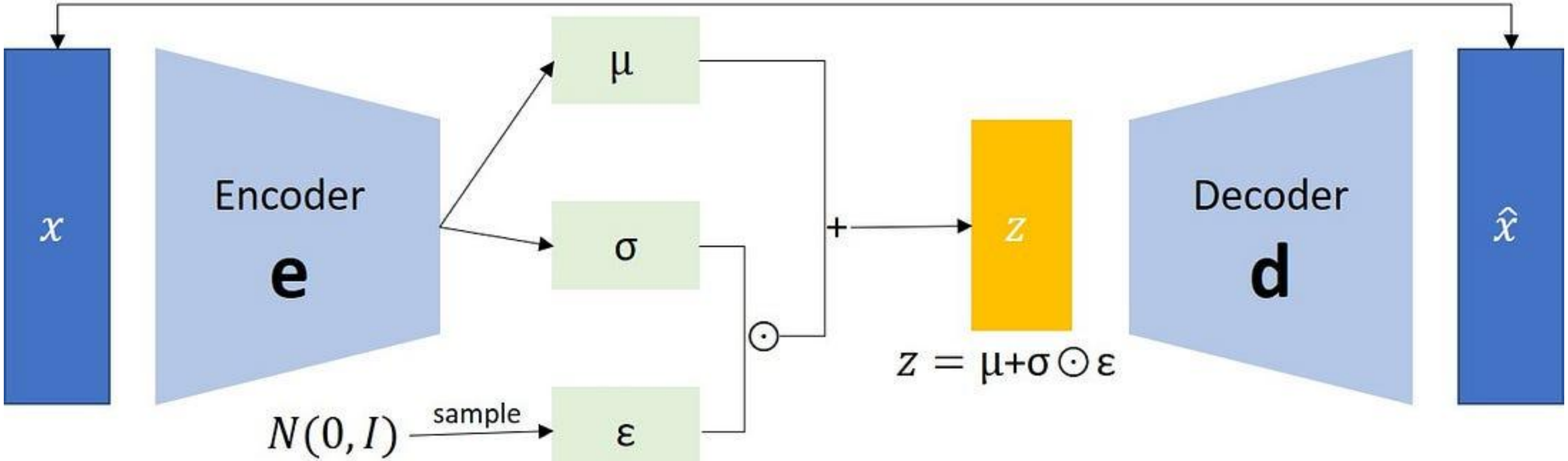
$$L = \sum q(z) \log \frac{p(x,z)}{q(z)} = \sum q(z) \log \frac{p(x|z)\,p(z)}{q(z)}$$

$$= \sum q(z) \left(\log p(x|z) + \log \frac{p(z)}{q(z)} \right)$$

$$= \underbrace{\sum q(z) \log p(x|z)}_{E_{q(z)}\left(\log p(x|z)\right)} + \underbrace{\sum q(z) \log \frac{p(z)}{q(z)}}_{-KL\left(q(z)\,||\,p(z)\right)}$$


Loss = $(x - \hat{x})^2$
x
Encoder
e
c
Decoder
d
$\hat{x}$
$e(x)$
$d(e(x))$

Minimize 1: $(x - \hat{x})^2$



$\varepsilon = \frac{x - \mu}{\delta}$

Minimize 2: $\frac{1}{2}\sum_{i=1}^{N}(\exp(\sigma_i) - (1+\sigma_i) + {\mu_i}^2)$

$$z = \mu + \sigma \odot \varepsilon$$

$$P(x) = \int_z P(z)P(x|z)dz$$

$$\text{Maximum } L = \sum_x \log P(x)$$

$$
\begin{aligned}
\log P(x) &= \int_z q(z|x) \log P(x) dz \\
&= \int_z q(z|x) \log \left(\frac{P(z,x)}{P(z|x)} \right) dz \\
&= \int_z q(z|x) \log \left(\frac{P(z,x)}{q(z|x)} \frac{q(z|x)}{P(z|x)} \right) dz \\
&= \int_z q(z|x) \log \left(\frac{P(z,x)}{q(z|x)} \right) dz + \int_z q(z|x) \log \left(\frac{q(z|x)}{P(z|x)} \right) dz \\
&= \int_z q(z|x) \log \left(\frac{P(z,x)}{q(z|x)} \right) dz + KL\left(q(z|x) || P(z|x) \right)
\end{aligned}
$$

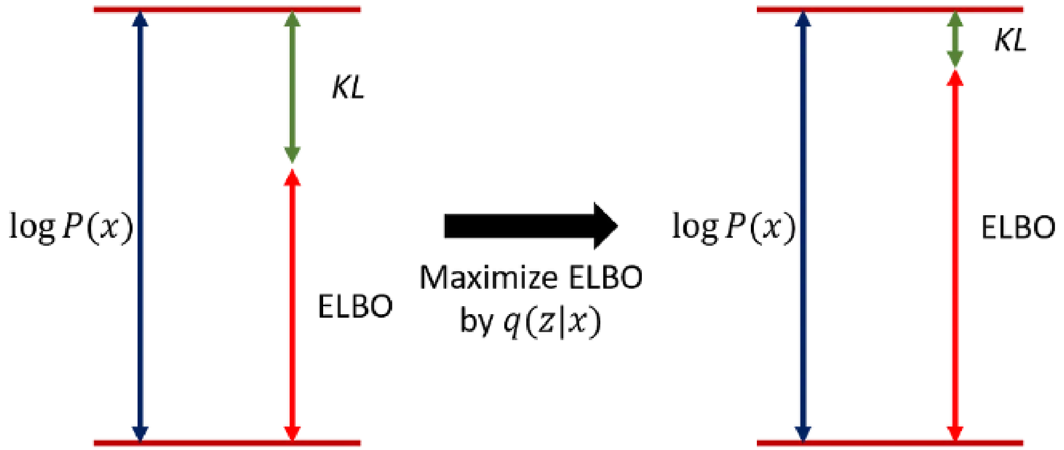
$$\log P(x) \geq \int_z q(z|x) \log \left(\frac{P(x|z)P(z)}{q(z|x)} \right) dz$$

We denote this lower bound as ELBO:

$$\log P(x) \geq \text{ELBO} = \int_z q(z|x) \log \left(\frac{P(x|z)P(z)}{q(z|x)} \right) dz = \text{E}_{q(z|x)} \left[\log \frac{p(x,z)}{q(z|x)} \right]$$

So we can revise the original form as:

$$\log P(x) = \text{ELBO} + KL\left(q(z|x) || P(z|x)\right)$$



$$
\begin{aligned}
\text{ELBO} &= \int_z q(z|x)\log\left(\frac{P(z,x)}{q(z|x)}\right)dz \\
&= \int_z q(z|x)\log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz \\
&= \int_z q(z|x)\log\left(\frac{P(z)}{q(z|x)}\right)dz + \int_z q(z|x)\log P(x|z)dz \\
&= -KL\left(q(z|x)||P(z)\right) + \int_z q(z|x)\log P(x|z)dz
\end{aligned}
$$

$$\text{Maximum} \int_z q(z|x) \log P(x|z)\, dz$$

$$\text{Maximum } \mathrm{E}_{q(z|x)} [\log P(x|z)]$$

$$\mathcal{N}(\mu, \delta)$$

$$= \frac{1}{\sqrt{2\pi}\,\delta} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\delta}\right)$$

$$\int_z q(z|x)\log P(z)\, dz = \int_z N\left(z;\mu,\sigma^2\right)\log N\left(z;0,I\right) dz$$

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

$$\int_z q(z|x)\log q\left(z|x\right) dz = \int_z N\left(z;\mu,\sigma^2\right)\log N\left(z;\mu,\sigma^2\right) dz$$

$$= -\frac{J}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{i=1}^{J}\left(1+\log\sigma_i^2\right)$$

$$\mathcal{N}(\mu, \Sigma)$$

$$\begin{bmatrix} \delta^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \delta^2 \end{bmatrix}$$

Then, we can write **-KL(*q*(*z*|*x*)||*P*(*z*))** as:

$$-KL\left(q(z|x)||P(z)\right) = \int_z q(z|x)\left(\log P(z) - \log q\left(z|x\right)\right) dz$$

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