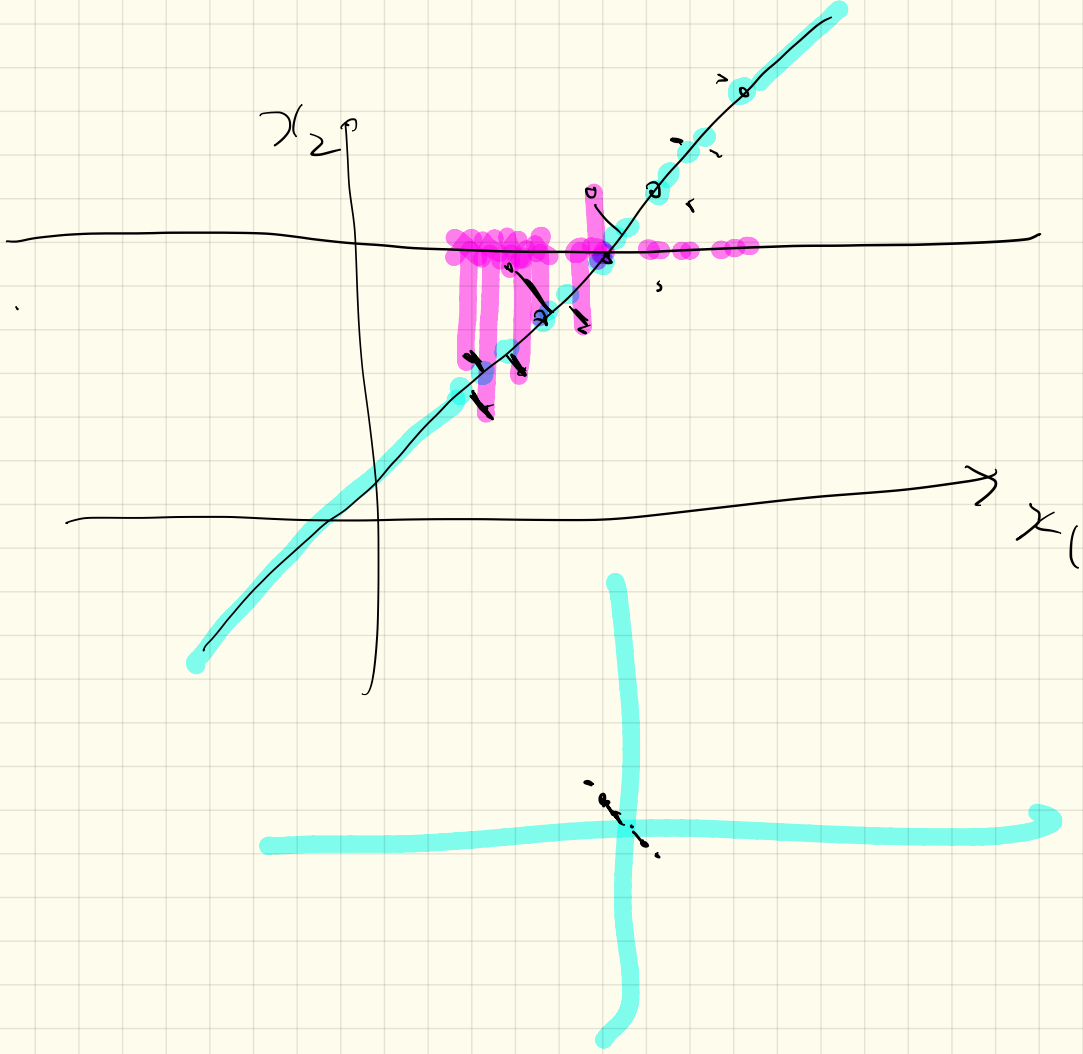
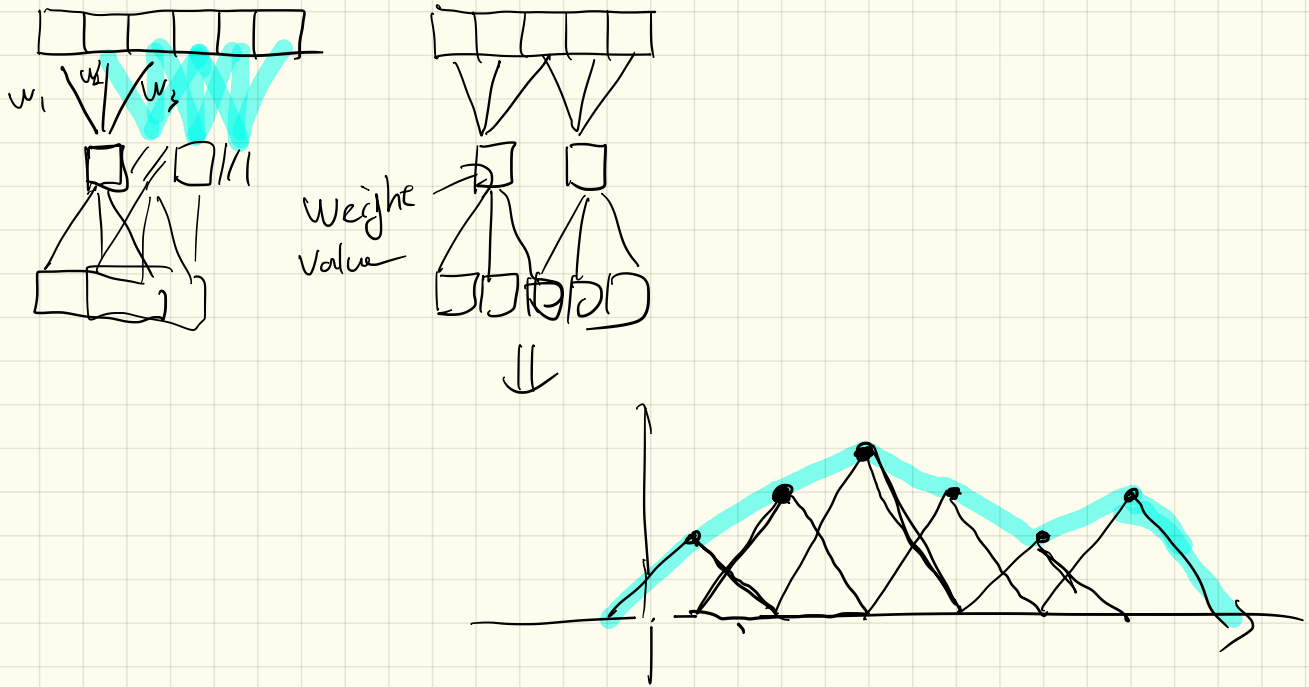


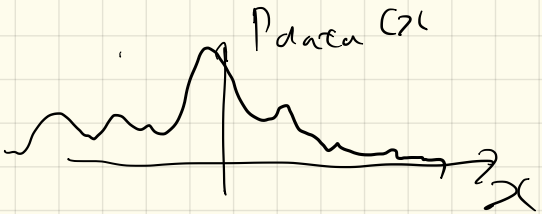

w_1
w_2
w_3
Weight
Value

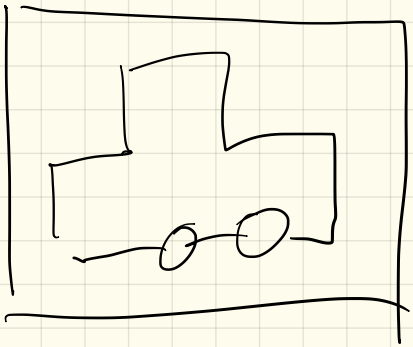
Generative model?

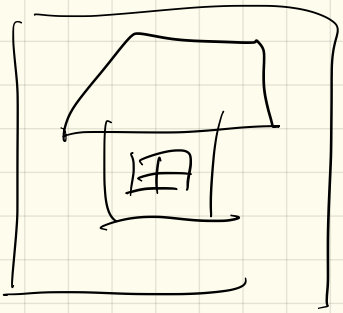
$P_{data}(x)$

$\Rightarrow$ Learn $P_{data}(x)$

$\Rightarrow$ Sample from the dist.



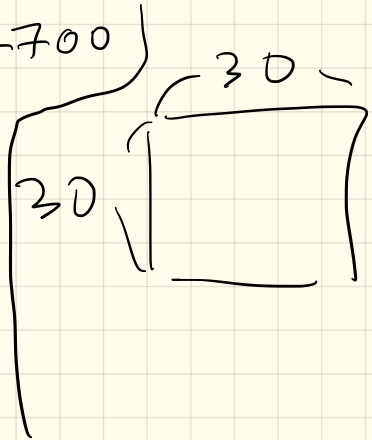




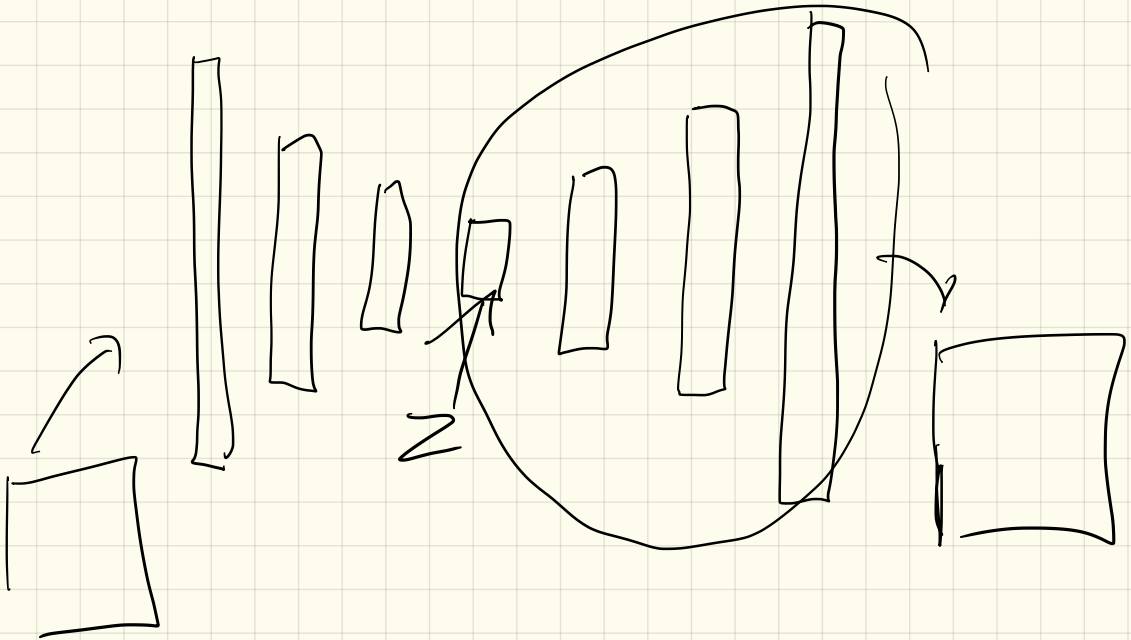
$N(\mu, \Sigma)$

$\mu \in \mathbb{R}^{2700}$

$\Sigma \in \mathbb{R}^{2700 \times 2700}$



$30 \times 30 \times 3$



$$l\left(P_{data}(x), \int P_\theta(x|z) P_\theta(z)\, dz\right) \Longrightarrow \frac{P_{data}(x)}{\int P_\theta(x|z)\, P_\theta(z)\, dz}$$

KL - divergence (Kullback-Leibler)

$$KL(P\|Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

$$= \int P(x) \log \frac{P(x)}{Q(x)} \, dx$$

$$= \mathbb{E}_{x \sim P(x)} \left[\log \frac{P(x)}{Q(x)} \right]$$

$KL(P\|Q) \neq KL(Q\|P)$

- $KL(P\|Q) \geq 0$

  $KL(P\|Q) = 0$; iff $P(x) = Q(x) \quad \forall x$

$KL(P \| Q)$

$= \sum_{x \in X} P(x) \log P(x)$ — Entropy of P

$- \sum_{x \in X} P(x) \log Q(x)$ ⇒ Cross entropy

in classification examples tasks

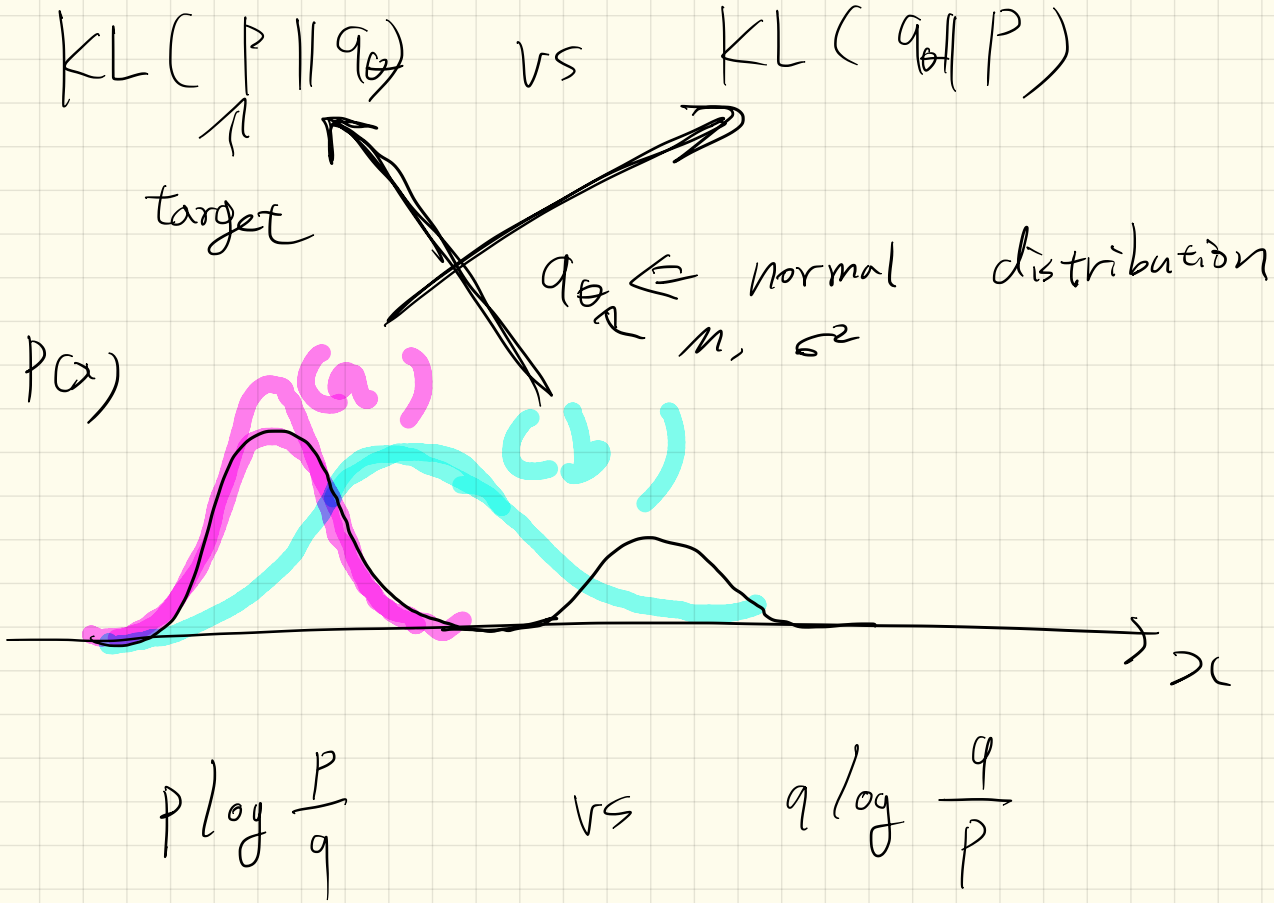
$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$KL(P_{data} \| P_\theta)$$

$$= \boxed{-\text{ entropy of } P_{data}}$$

$$+ \text{ Cross entropy of } P_{data} \text{ and } P_\theta$$

$\Rightarrow$ maximizing log likelihood.


KL(P || q_θ) vs KL(q_θ || P)
target
q_θ ⇐ normal distribution
μ, σ^2
P(x)
(a)
(b)
x
$p \log \frac{p}{q}$ vs $q \log \frac{q}{p}$

$$0 \log 0 = 0$$

$$\lim_{x \to 0} x \log x = 0$$