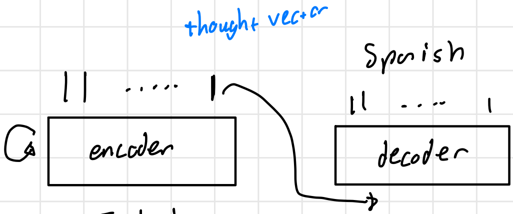
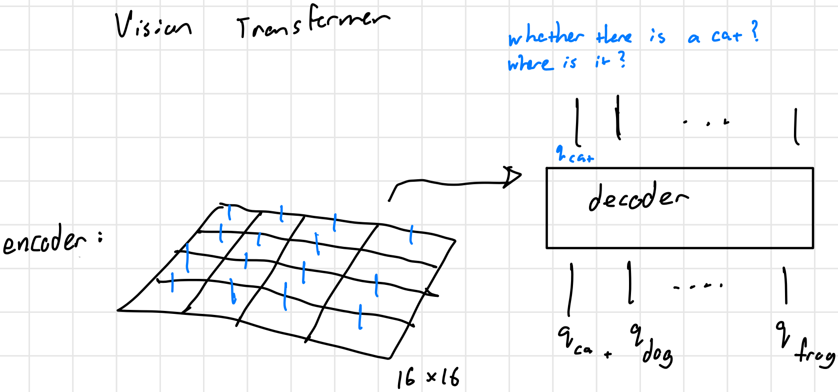
Lecture 19

- Neural Network ≡ team of vectors.

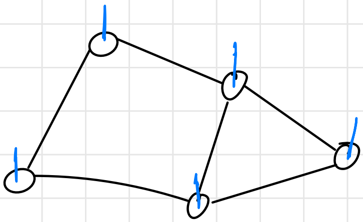
Transform



Vision Transformer

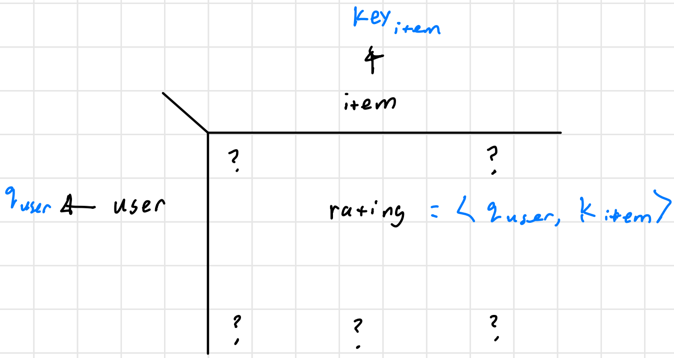


- Grap Neural Network (GNN)
  - social network



thought vector
node embedding

- Recommender System :



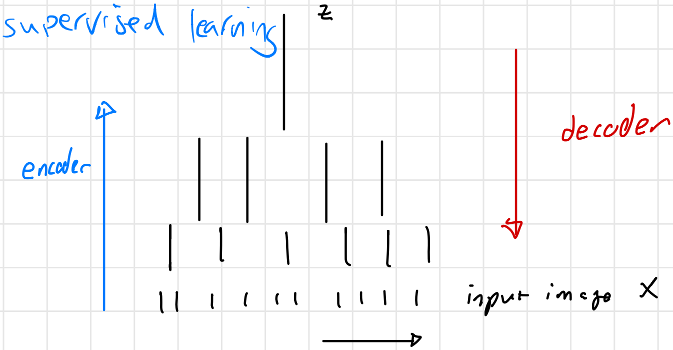
- Generative Model: unsupervised learning

z: latent vector

encoder ↑ ↓ decoder

X: image

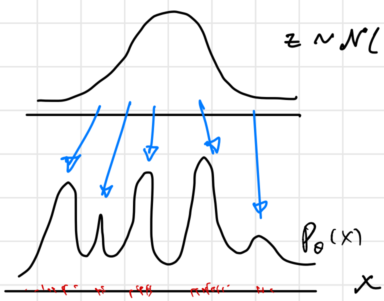
Auto-encoder



- Generator Model (decoder)

$$z \sim \mathcal{N}(0, I_d)$$

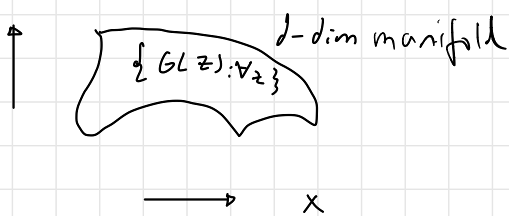
$$X = G(z) + \varepsilon$$



density estimation

observed examples sparse in high dimensional space

- Manifold Principle:



- GAN (Generative Adversarial Networks)

  - Discriminator : $p(x) = P(x \text{ is real}) = P(Y=1 \mid x)$

Real	$1, \ldots, i, \ldots, n$	x_i	$y_i = 1$
Fake	$1, \ldots, i, \ldots, n$	$\tilde{x}_i = G(\tilde{z}_i)$	$\tilde{y}_i = 0$

- log likelihood :

$$\frac{1}{n}\sum_{i=1}^{n} \log p(y_i \mid x_i) + \frac{1}{n}\sum_{i=1}^{n} \log(\tilde{y}_i \mid \tilde{x}_i)$$

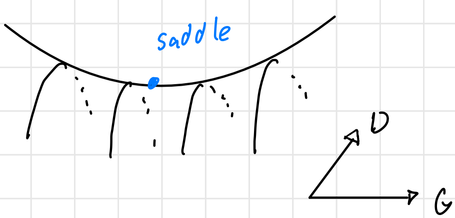
$$= \frac{1}{n}\sum_{i=1}^{n} \log D(x_i) + \frac{1}{n}\sum_{i=1}^{n} \log(1 - D(\tilde{x}_i))$$

$\tilde{x}_i \downarrow G(\tilde{z}_i)$

- Value :

$$V(G, D) = \frac{1}{n}\sum_{i=1}^{n} \log D(x_i) + \frac{1}{n}\sum \log(1 - D(G(\tilde{z}_i)))$$

$$\min_G \max_D V$$ Adverserial Game



- VAE (Variational auto-encoder)

vector code

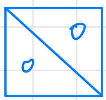
$z \sim \mathcal{N}(0, I_d)$ — $\dim(z)$

encoder $q_\varphi(z|x)$

decoder $p_\theta(x|z) \sim \mathcal{N}(g_\theta(z), \sigma^2 I_d)$

x

$\mathcal{N}(\mu_\varphi(x), V_\varphi(x))$



assumed diagonal

So $z = \mu_\varphi(x) + V_\varphi^{1/2}(x) \cdot e$, $e \sim \mathcal{N}(0, I_d)$

$p_\theta(x, z) = p(z)\, p_\theta(x|z)$ factorization

$p_\theta(x) = \int p_\theta(x, z)\, dz$ marginalization (intractable)

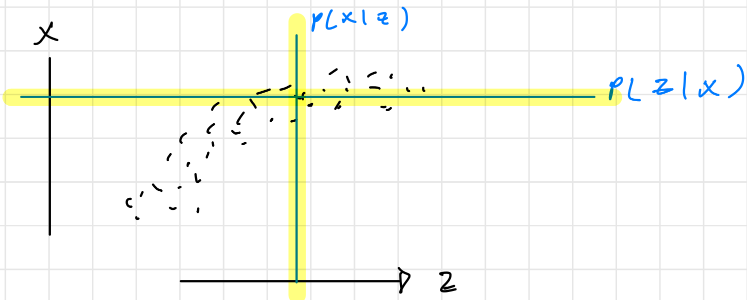
$p_\theta(z|x) = \dfrac{p_\theta(x, z)}{p_\theta(x)}$ conditioning (intractable)

$q_\varphi(z|x)$ serves as tractable approx to $p_\theta(z|x)$

z

tractable $q_\varphi(z|x)$ $p_\theta(x|z)$ intractable

x



- Maximum Liklihood

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log \boxed{p_\theta(x_i)}$$

intractable

density estimation

- Translation :
  - If $n \to \infty$

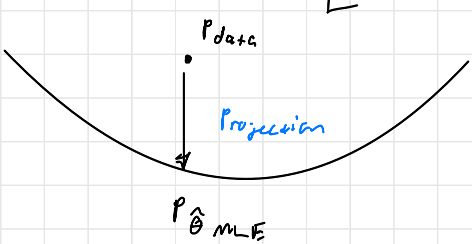
$$L(\theta) \doteq \mathbb{E}_{p_{data}(x)}\left[\log p_\theta(x)\right]$$

$x_1 \cdots x_i \cdots x_n \overset{iid}{\sim} p_{data}(x)$

$$\max_\theta L(\theta) = \min_\theta \text{ const} - L(\theta)$$

$$\mathbb{E}_{p_{data}}\left[\log p_{data}(x) - \log p_\theta(x)\right]$$

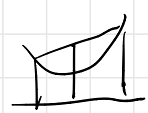
$$= \mathbb{E}_{p_{data}}\left[\log \frac{p_{data}(x)}{p_\theta(x)}\right] = D_{KL}\left(p_{data}(x) \,\|\, p_\theta(x)\right)$$



$\{ p_\theta, \forall\theta \}$

$$D_{KL}(p\|q) = E_p\left[\log\frac{p(x)}{q(x)}\right] = E_p\left[-\log\frac{q(x)}{p(x)}\right] \geq -\log\left(E_p\left[\frac{q(x)}{p(x)}\right]\right) = 0$$

$$\int \frac{q(x)}{p(x)} p(x)\,dx = \int q(x)\,dx = 1$$



- MLE:

$$\min_{\theta} D_{KL}\big(p_{data}(x) \,|\, \underbrace{p_\theta(x)}_{\text{Intractable}}\big)$$

VAE:

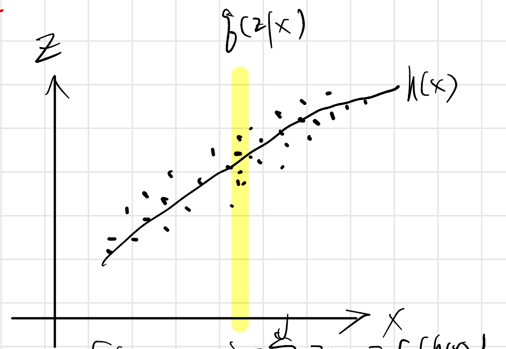
$$\min_{\theta,\varphi} D_{KL}\big(p_{data}(x, z) \,|\, \underbrace{p_\theta(x, z)}_{\text{Tractable}}\big)$$

$p_\theta(x,z) = p(z)\underbrace{p_\theta(x|z)}_{\text{decoder}}$

$p_{data}(x,z)$ → (z: ?)

$$p_{data}(x) \cdot \underbrace{q_\varphi(z|x)}_{\text{encoder}} \quad \text{tractable}$$

- Understanding:



$E(z) = \frac{1}{N}\sum_{i=1}^{N} z_i = E(h(x)) = E_x E_{z|x}(z)$

$E(z|x) = h(x)$

↓ $h(x_i)$

does not change group sum

$$D_{KL}\big(p_{data}(x,z) \,|\, p_\theta(x,z)\big)$$

$$= \mathbb{E}_{p_{data}}\left[\log \frac{p_{data}(x,z)}{p_\theta(x,z)}\right]$$

$$= \mathbb{E}_{p_{data}}\left[\underbrace{\log \frac{p_{data}(x)}{p_\theta(x)}}_{\text{MLE}} + \underbrace{\log \frac{q_\varphi(z|x)}{p_\theta(z|x)}}_{\text{Extra}}\right]$$

$$= \mathbb{E}_{p_{data}}\left[\log p_{data}(x) - \log p_\theta(x) + \log \frac{q_\varphi(z|x)}{p_\theta(z|x)}\right]$$

↓

$$= \mathbb{E}_{p_{data}(x)} \mathbb{E}_{q_\varphi(z|x)}\left[\quad\right]$$

$$= \mathbb{E}_{p_{data}(x)}\left[\log p_{data}(x) - \log p_\theta(x) + \mathbb{E}_{q_\varphi(z|x)}\left[\log \frac{q_\varphi(z|x)}{p_\theta(z|x)}\right]\right]$$

$$= \mathbb{E}_{p_{data}(x)}\left[\log p_{data}(x) \left\{\log p_\theta(x) - D_{KL}(q_\varphi(z|x) \| p_\theta(z|x))\right)\right]$$

Intractable

- Approx $\log p_\theta(x)$ by

ELBO : $\log p_\theta(x) - D_{KL}(q_\varphi(z|x) \| p_\theta(z|x))$

Tractable

Evidence Lower Bound

Problem : (1) $p_{data}(x)\, q_\phi(z|x) \xrightarrow{\int dx} q(z)$ not $N(0, I)$

(2) $p_\theta(x|z) \sim N(g_\theta(x), \sigma^2 I_\theta)$ not accurate

($\sigma^2 I_\theta$ ↓ big)

• Diffusion : more careful VAE   no φ to learn

Encoder

$q(z|x_0)$ add noise

- clean image

$x = x_0 \rightarrow$ [$x_1 \rightarrow x_2 \rightarrow \dots x_{t-1} \rightarrow x_t \rightarrow x_{t+1} \rightarrow \dots \rightarrow x_T$] $\sim N(0, \ )$ accurate (1)

z

$x_t = x_{t-1} + N(0, \sigma^2 I)$

Supervised learning $p(x_{t-1}|x_t) \sim N(x_t + \sigma \nabla_x \log p_t(x_t), \sigma^2 I)$ accurate (2)

Decoder : $p(z)\, p(x|z)$

When noise is small