1 Generative models

$p(\bar{x})$

$D = \{\bar{x}\}$

$p_{model}(\bar{x}|\bar{\theta}) \rightarrow p_{data}(\bar{x})$

$$p(D|\bar{\theta}) = \prod_n p(\bar{x}_n|\bar{\theta}) \xrightarrow[\theta]{} \max$$

$$p_{data}(\bar{x}) = \text{Unif}(D) = \frac{1}{N}\sum_n \delta(\bar{x}-\bar{x}_n)$$

$$KL(p_{data}\,\|\,p_{model}) = \int p_{data}(\bar{x}) \log \frac{p_{data}(\bar{x})}{p_{model}(\bar{x})} d\bar{x} =$$

$\xrightarrow[\bar{\theta}]{} \min$

$$= H(p_{data}) - \int p_{data} \cdot \log p_{model}\, d\bar{x} =$$

$$= \text{Const} - \frac{1}{N}\sum_n \log p_{model}(\bar{x}_n) \xrightarrow[\bar{\theta}]{} \max$$

Naive Bayes – generative

$$p(\bar{x}, y) = p(y) \cdot \prod_{i=1}^{d} p(x_i|y)$$

$\theta_y \qquad \theta_y$

$$\log p(\bar{x}, y) = \log p(y) + \sum_i \log p(x_i|y)$$

$$\theta_k = \log p(y = C_k)$$

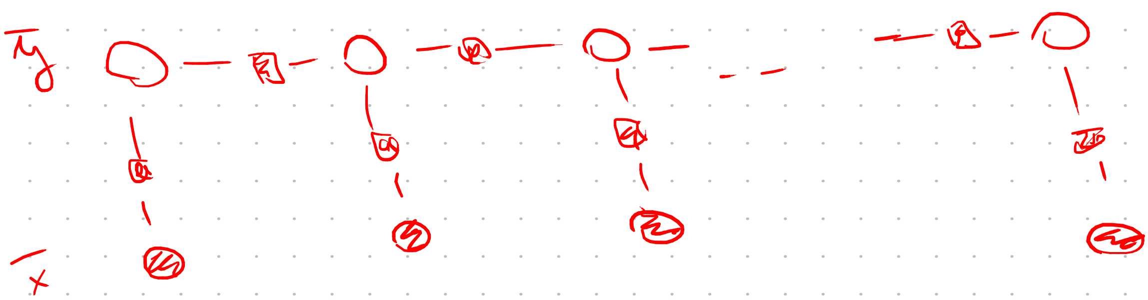
$$\theta_{ink} = \log p(x_i = w_n \,|\, y = C_k)$$

$$\log p(D|\bar{\theta}) = \sum_d \sum_k \theta_k [y_d = C_k] + \sum_d \sum_k \sum_n \sum_i \theta_{ink} [y_d = C_k][x_{id} = w_n]$$

Log. regr

$$p(y = C_k|\bar{x}) \propto e^{\bar{w}_k \bar{x}}$$

generative – discriminative pair



$p(\bar{y}|\bar{x})$

$\bar{y}$

2 Deep Generative Models $p(\bar{x})$

GAN

Implicit density

$\bar{z} \to \square \to \bar{x} \sim p(\bar{x})$

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

$p(\bar{x}) = \ldots$

Explicit density

VAE

Approximate density

$p(\bar{x}) \approx q(\bar{x}) = \prod q_i(\bar{x})$

$KL(q \| p) \to \min$

Tractable density

$p(\bar{x}) = ?$

Flow-based models

$p(\bar{x}) = f(\bar{z} = \bar{x}_0) =$

$= f_m \circ f_{m-1} \circ \ldots \circ f_1(\bar{x}_0)$

MAF, IAF, —

diffusion-based models

Autoregressive models

$p(\bar{x}) = p(x_1)\, p(x_2|x_1)\, p(x_3|x_1, x_2) \ldots p(x_n|x_1 \ldots x_{n-1})$

Naive $\quad p(\bar{x}, y) = p(y) \cdot p(x_1|y) \ldots p(x_n|y)$

HMM $\quad p(\bar{x}, \bar{y}) = p(y_1)\, p(x_1|y_1)\, p(y_2|y_1)\, p(x_2|y_2) \ldots$

RNN $\quad p(\bar{y}, \bar{h} | \bar{x}) = p(h_0)\, p(h_1, y_1 | h_0, x_1)\, p(h_2, y_2 | h_1, x_2) \ldots$

y_1

$h_0 - \square - h_1 - \ldots$

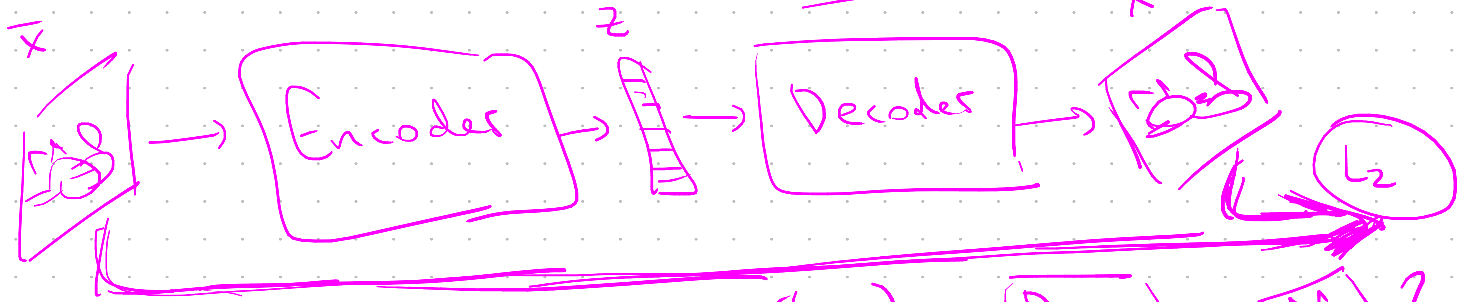
x_1

LLM: $p(x_t | x_1 \ldots x_{t-1})$

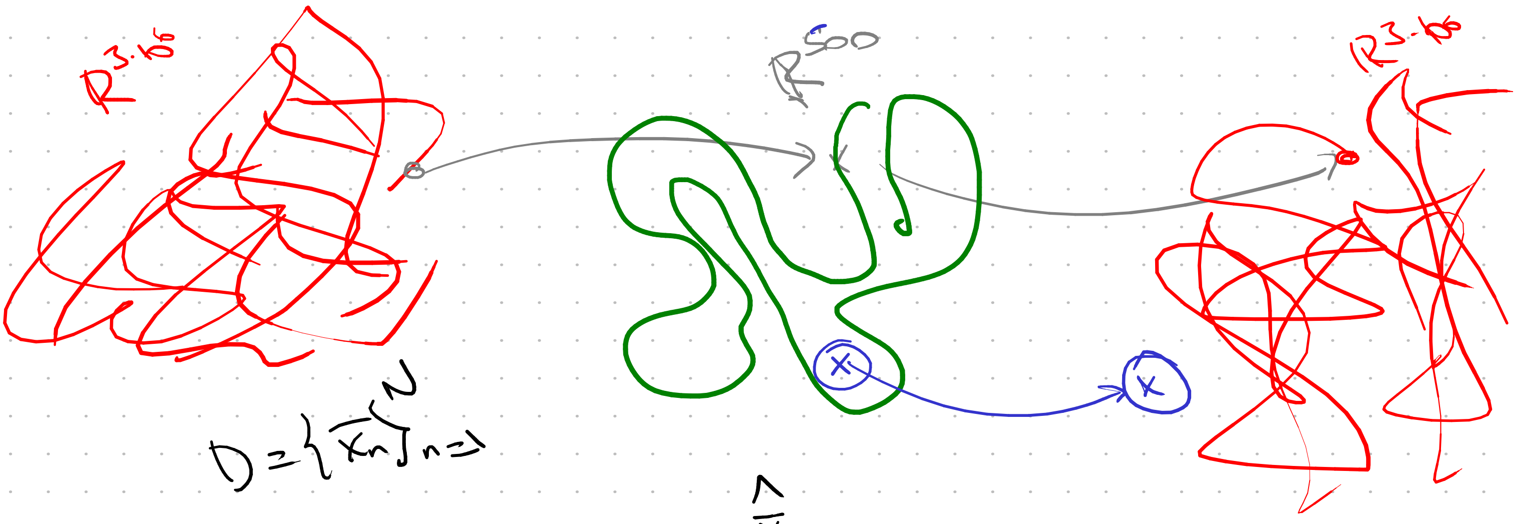
3 Generative Adversarial Networks (GAN)

Autoencoders

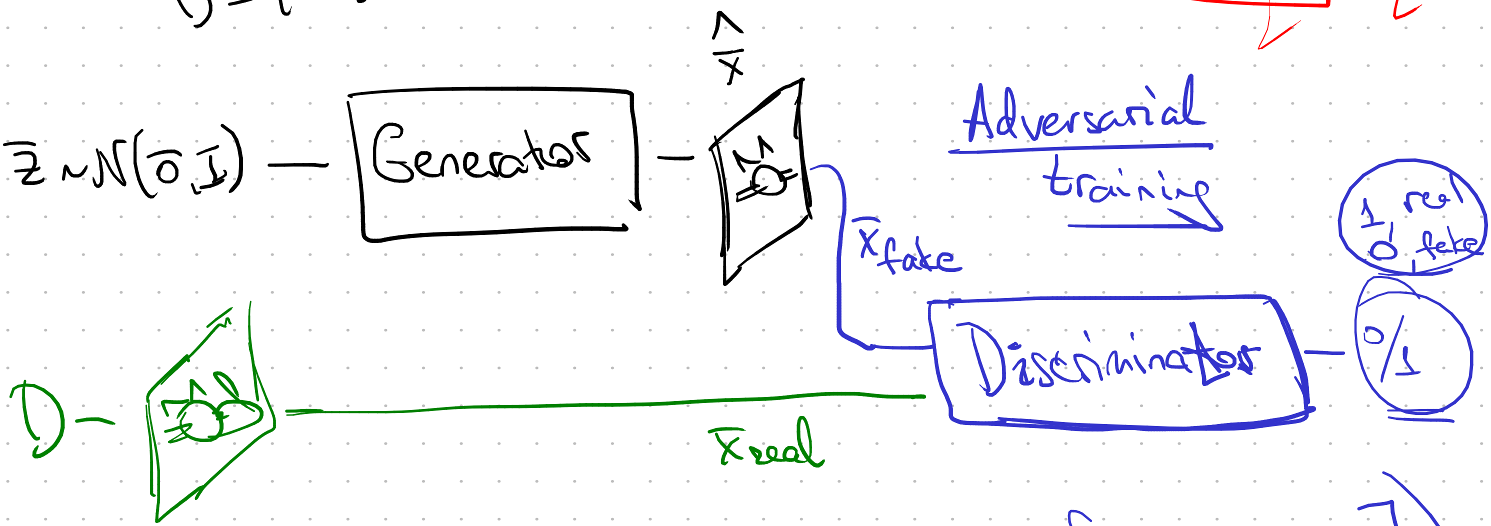
$D = \{\bar{x}\}$



$\bar{z} \sim \mathcal{N}(\bar{0}, I) \to \text{Dec} \to$?



$D = \{\bar{x}_n\}_{n=1}^{N}$



$$L_D = -\left(\mathbb{E}_{\bar{x}_{real} \sim D}\left[\log D(\bar{x})\right] + \mathbb{E}_{\bar{x}_{fake} \sim p_g(\bar{x})}\left[\log(1 - D(\bar{x}))\right] \right) \quad \xrightarrow[\bar{\theta}_D]{} \min$$

$$L_G = \mathbb{E}_{\bar{z} \sim p(\bar{z})}\left[\log(1 - D(G(\bar{z})))\right] \quad \xrightarrow[\bar{\theta}_G]{} \min$$

- fix θ_G, train θ_D
- fix θ_D, train θ_G

$$\underbrace{\min_{\bar{\theta}_G}}\ \underbrace{\max_{\bar{\theta}_D} L(\bar{\theta}_D, \bar{\theta}_G)}$$

$p_G(\bar{x})$

$$\underbrace{D^*_G(\bar{x})} = \frac{p_{data}(\bar{x})}{p_{data}(\bar{x}) + p_G(\bar{x})} \Rightarrow G^*: p_{data} = p_g$$

$$\min_{\bar{\theta}_G} L(\bar{\theta}_D^*, \bar{\theta}_G) = JSD(p_{data}, p_g) = KL\left(p_{data} \,\middle\|\, \frac{p_{data} + p_g}{2}\right) + KL\left(p_g \,\middle\|\, \frac{p_{data} + p_g}{2}\right)$$