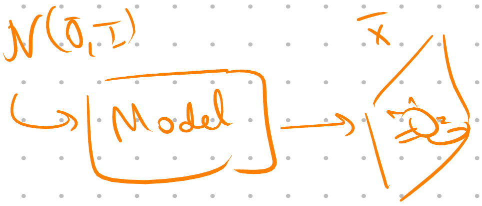
$\bar{x}$ $p(\bar{x})$

Generative models

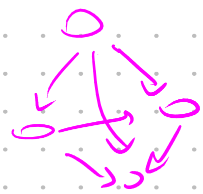
Explicit density

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


Implicit density




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



$$p(\bar{x}|\bar{\theta}) = p(x_1, x_2, \dots, x_d|\bar{\theta}) =$$
$$= p(x_1|\bar{\theta})\, p(x_2|x_1, \bar{\theta})\, p(x_3|x_1, x_2, \bar{\theta}) \dots p(x_d|x_1, \dots, x_{d-1}, \theta)$$

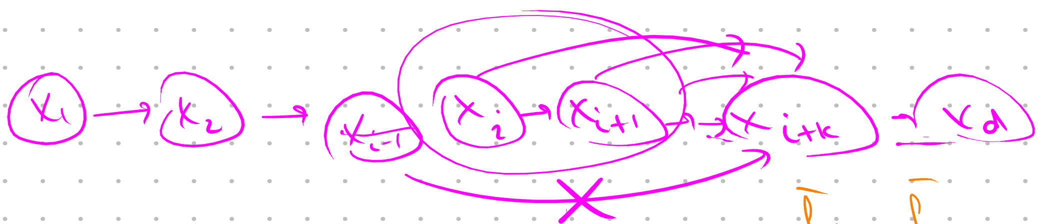
$p(x_i|\bar{x}_{1..i-1}, \bar{\theta})$ — autoregressive model

Sliding window

$$p(x_i|x_1, \dots, x_{i-1}, \bar{\theta}) =$$
$$= p(x_i|x_{i-1}, x_{i-2}, \dots, x_{i-k}, \bar{\theta})$$

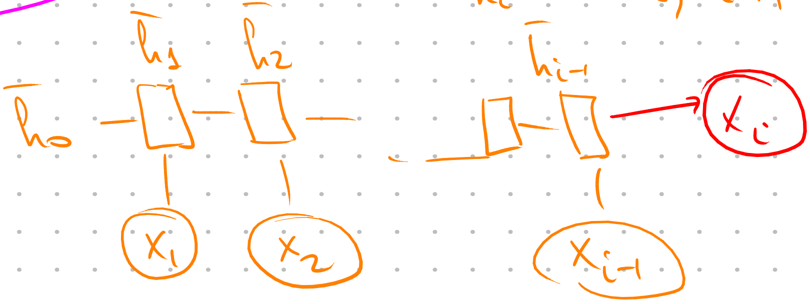
Hidden state

$$p(x_i|x_1 \dots x_{i-1}, \bar{\theta}) =$$
$$= p(x_i|\bar{h}_{i-1}, \bar{\theta}),$$




rge

$$\bar{h}_i = f(x_i, \bar{h}_{i-1}, \bar{\theta})$$



## Explicit density

**Tractable density**: $p(\bar{x}|\bar{\theta}) = \dots$

- Autoregressive tractable density
- Flow-based models

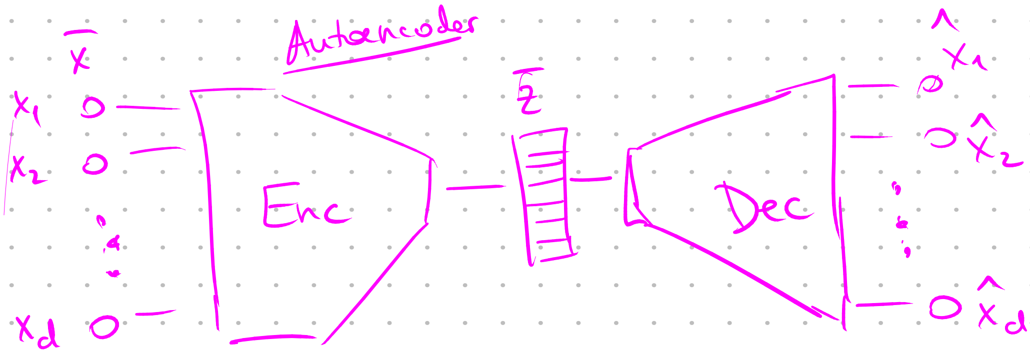
**Approximate density**: $p(\bar{x}) \approx \boxed{q(\bar{x}|\bar{\theta})}$, $KL(q\|p) \to \min$

- Variational appr. — VAE
- MCMC

---

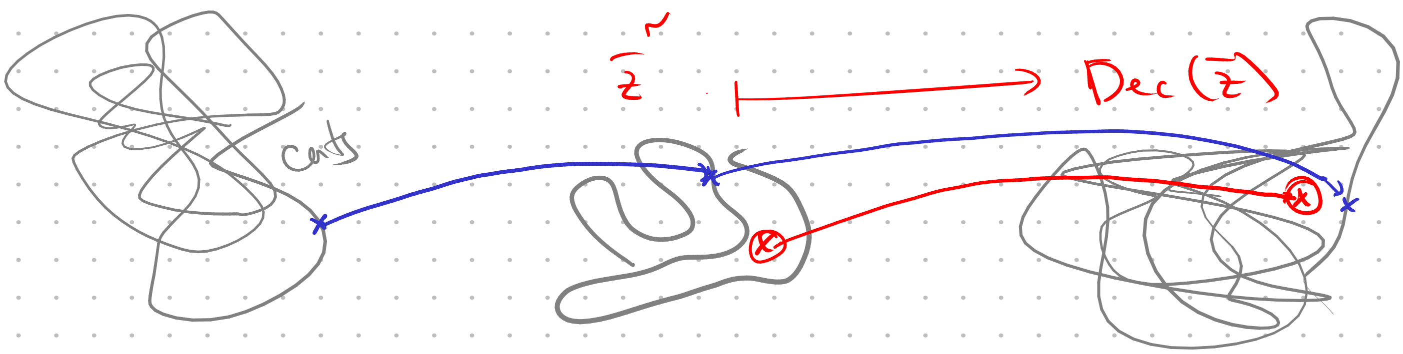
## MADE – masked autoencoders for density estimation

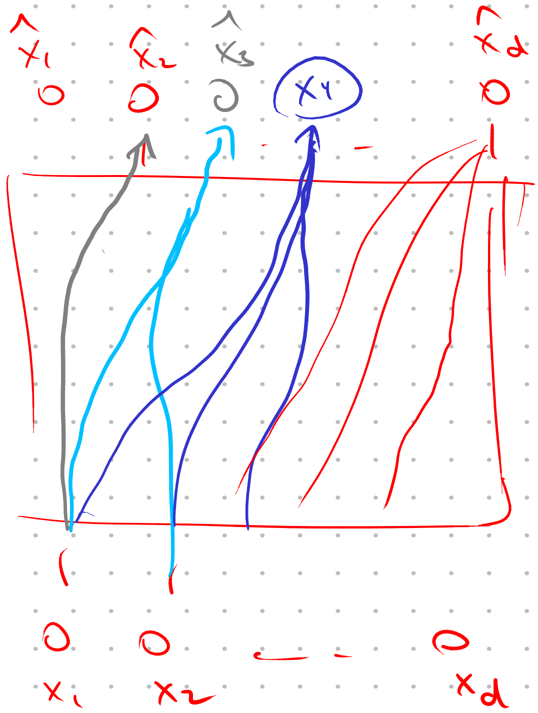
Autoencoder




$$L = \|\hat{x} - \bar{x}\| \to \min$$

$$\mathbb{R}^{3\cdot 10^6} \xrightarrow{\text{Enc}} \mathbb{R}^{500} \xrightarrow{\text{Dec}} \mathbb{R}^{3\cdot 10^6}$$





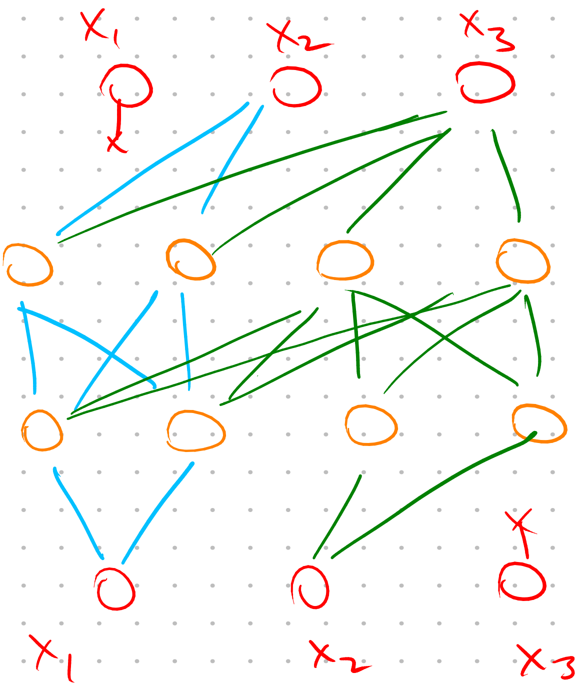

$x_1 \sim p(x_1)$

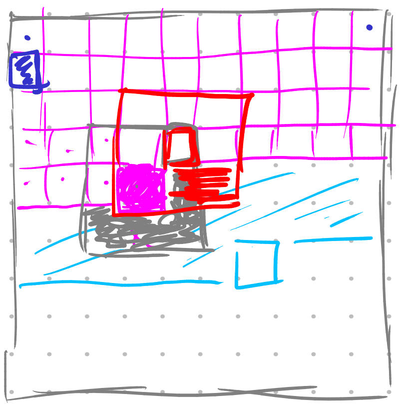
$x_2 \sim p(x_2|x_1)$

$x_3 \sim p(x_2|x_1, x_2)$

$x_d \sim p(x_d|x_1 \ldots x_{d-1})$

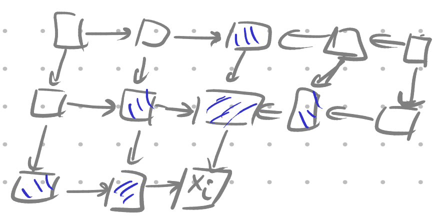
## MADE

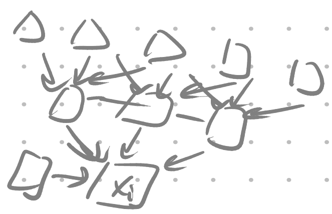


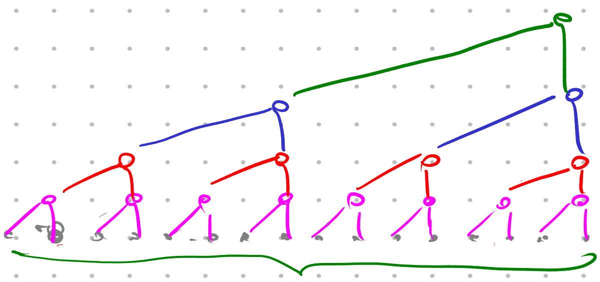



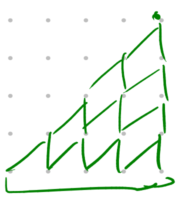
Pixel RNN

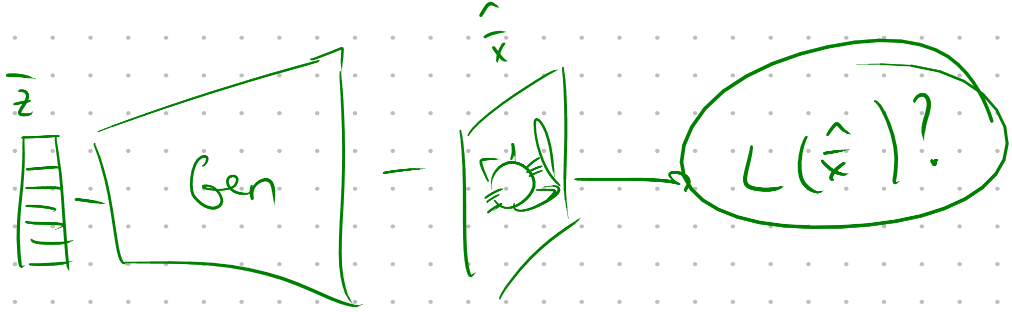
Pixel CNN



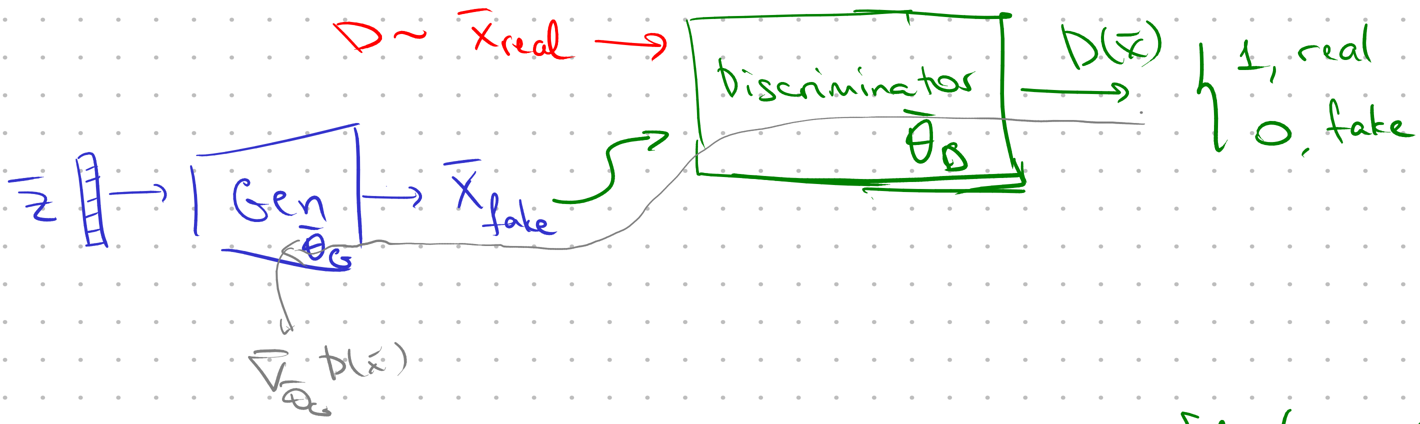








## Generative adversarial networks




$$L_D(\bar\theta_D) = \underbrace{\mathbb{E}_{\bar x_{real}\sim D}\left[\log D(\bar x)\right]} + \mathbb{E}_{\bar x_{fake}\sim p_G(\bar x)}\left[\log(1-D(\bar x))\right] \xrightarrow[\bar\theta_D]{} \max$$

$\bar\theta_G \searrow \min$

$$L_G(\bar\theta_G) = \mathbb{E}_{\bar x_{fake}}\left[\log(1-D(\bar x))\right] =$$

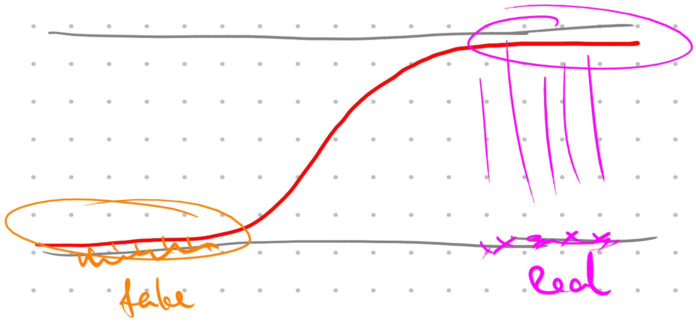
$$= \mathbb{E}_{\bar z\sim p(\bar z)}\left[\log(1-D(G(\bar z)))\right] \xrightarrow[\bar\theta_G]{} \min$$

$$\boxed{\min_{\bar\theta_G}\left[\max_{\bar\theta_D} L(\bar\theta_D, \bar\theta_G)\right]}$$

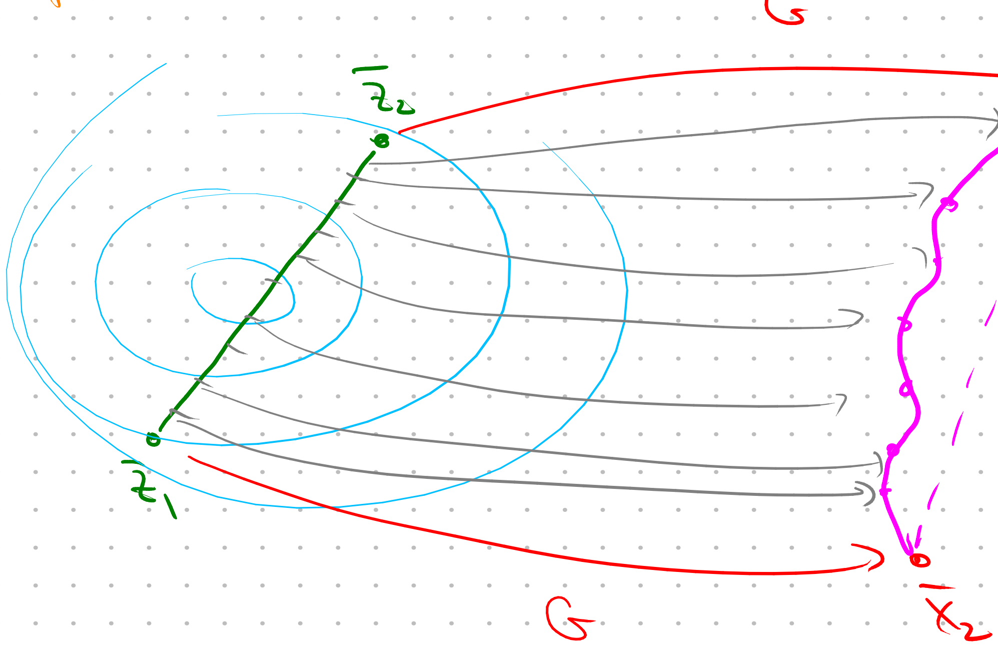
- fix G, train D
- fix D, train G

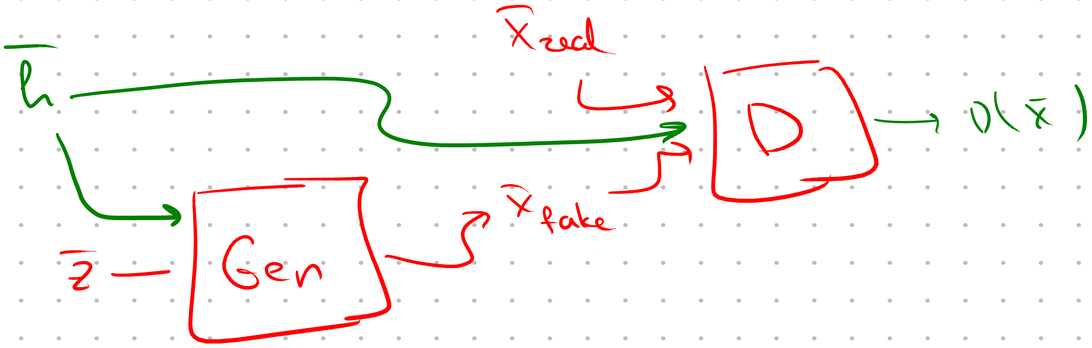
fix G

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

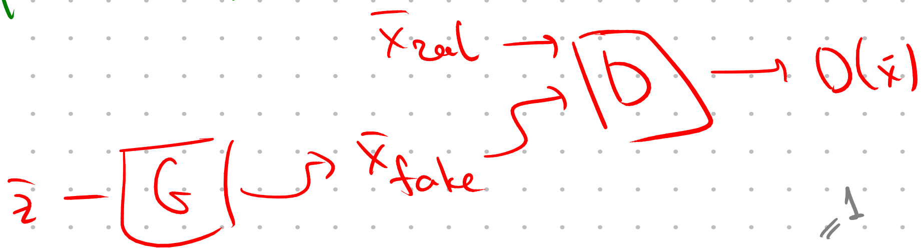


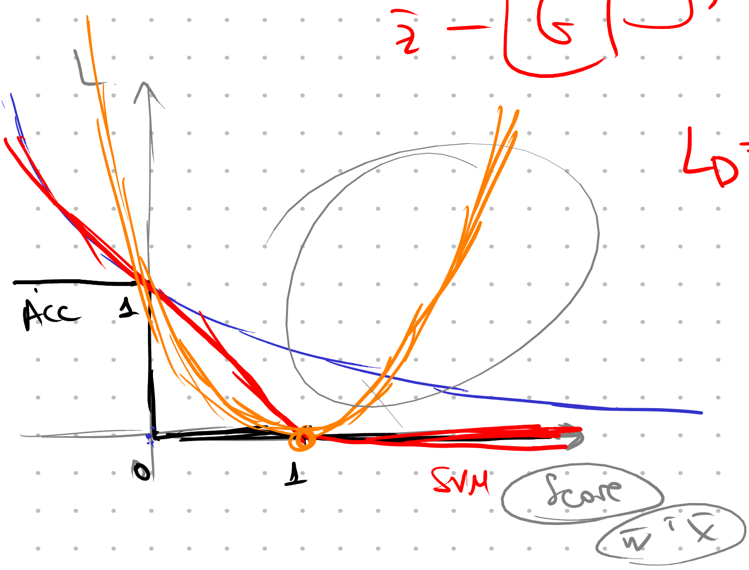

$$-\mathbb{E}_{\bar{z}\sim p(\bar{z})}\left[\log D(G(\bar{z}))\right]$$




① Least Squares GAN




$$L_D = \mathbb{E}_{p_{data}}\left[(D(\bar{x}) - \overset{1}{\overset{\shortparallel}{b}})^2\right] + \mathbb{E}_{p_G}\left[(D(\bar{x}) - \overset{0}{\overset{\shortparallel}{a}})^2\right]$$

$1 =$ $= -1$

$$L_G = \mathbb{E}_{\bar{z}}\left[(D(G(\bar{z})) - \underset{0}{\underset{\shortparallel}{\overset{1}{\overset{\shortparallel}{c}}}})^2\right]$$