Generative Adversarial Networks

Setup: Given $\mathcal{D} := \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^d$, $\underline{x_i \sim q(x)}$ empirical data distributi

$$p_\theta(x) = \int p(x,z)\,dz = \int \underbrace{p(x|z)}_{\mathcal{N}(x|\mu_\theta(z), \Sigma_\theta(z))} \underbrace{p(z)}_{\mathcal{N}(0,I)}\,dz \quad , \quad p(z|x) \approx q_\phi(z$$

ELBO

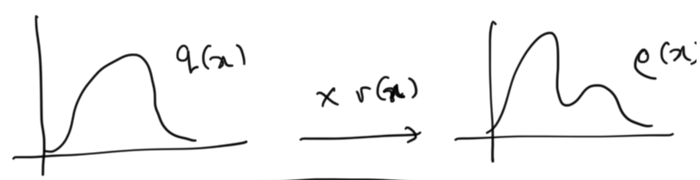
$$\mathcal{L}(\theta,\phi) = \underbrace{-\log p_\theta(x)}_{\text{MLE}} \leq \mathbb{KL}[q_\phi(z|x) \,\|\, p(z)] - \mathbb{E}_{z \sim q_\phi(z|x)}[\log p_\theta(x|z)]$$

$$\Updownarrow$$

$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathbb{KL}[q(x) \,\|\, p_\theta(x)] = \mathbb{E}_{x \sim q(x)}\left[\log \frac{q(x)}{p_\theta(x)}\right.$$

Density-ratio estimation:

$$r(x) = \frac{p(x)}{q(x)}$$



$x_i \sim p(x)$

$x_j \sim q(x)$

$$X = [\underbrace{x_1, \dots, x_n}_{p(x)}, \underbrace{x'_1, \dots, x'_n}_{q(x)}] \quad 2n \times d$$

$$y = [+1, \dots, +1, -1, \dots, -1] \quad 2n \times 1$$

$$\boxed{p(x) = p(x|y=+1)}, \quad \boxed{q(x) = p(x|y=-1)}$$

$$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$

$$r(x) = \frac{p(x)}{q(x)} = \frac{p(x|y=+1)}{p(x|y=-1)} \overset{\text{Bayes}}{=} \frac{\frac{p(y=+1|x)\,\cancel{p(x)}}{p(y=+1)}}{\frac{p(y=-1|x)\,\cancel{p(x)}}{p(y=-1)}}$$

$$= \cancel{\frac{p(y=-1)}{p(y=+1)}}^{\to 1} \cdot \frac{p(y=+1|x)}{p(y=-1|x)} = \frac{p(y=+1|x)}{1-p(y=+1|x)} \approx \frac{S_\theta(\cdot}{1-S_\theta(}$$

GANs :

Goal is to construct a model for $p_\theta(x)$.

$$G_\theta(z) = x \quad , \quad \underline{z \sim p(z)}^{\to \mathcal{N}(0,I)} , \; z \in \mathbb{R}^q , \quad x \in \mathbb{R}^d$$

$$G_\theta : \mathbb{R}^q \longrightarrow \mathbb{R}^d \quad , \quad x \sim p_\theta(x)$$

$\left\{ \begin{array}{l} \text{We want to train this generative mapping such that the distribu} \\ p_\theta(x) \text{ is as close as possible to } q(x) \end{array} \right.$ { empirical dist. of the observed data

Unlike MLE approaches, here we introduce a discriminator

$$D_\varphi : \mathbb{R}^d \longrightarrow [0,1] \quad , \quad D_\varphi(x)$$

Training :

$$\theta^*, \varphi^* = \underset{\theta}{\min}\,\underset{\varphi}{\max} \; \mathbb{E}_{x\sim q(x)}\left[\log D_\varphi(\overset{\text{true}}{x})\right] + \mathbb{E}_{z\sim p(z)}\left[1 - \log D_\varphi(\overbrace{\underbrace{G_\theta(z)}}^{\text{fake}})\right]$$

Algorithm :

~ for # iter .

~ for (k) steps do :

sample $z_1, \dots, z_m \sim p(z)$

Sample $x_1, \dots, x_m \sim q(x)$

keep θ params fixe

$$\varphi_{n+1} = \varphi_n + \eta \nabla_\varphi \left\{ \frac{1}{m} \sum_{i=1}^{m} \log(D_\varphi(x_i)) + \log[1 - D_\varphi(G_\theta(\dots\right.$$

while θ is kept fixed.

binary cross-entropy

for (t) steps do :

Sample $z_1, \dots, z_m \sim p(z)$

$$\theta_{n+1} = \theta_n - \eta \nabla_\theta \left\{ \frac{1}{m} \sum_{i=1}^{m} \log[1 - D_\varphi(G_\theta(z_i))] \right.$$

while φ is kept fixed.

Remarks :

1.) For $G_\theta(z)$ fixed, the optimal discriminator is :

$$\boxed{D^*_\varphi(x) = \frac{q(x)}{q(x) + p_\theta(x)}}$$

$$J(\theta) = \max_\varphi L(\theta, \varphi) = \mathbb{E}_{x \sim q(x)}\left[\log \frac{q(x)}{q(x) + p_\theta(x)}\right] + \mathbb{E}_{x \sim p_\theta(x)}\left[\log \frac{p_\theta(\dots}{q(x)+\dots}\right.$$

The global minimum of $J(\theta)$ can be achieved iff $p_\theta(x) = \dots$

$$\boxed{J(\theta^*) = -\log 4}$$

Goodfellow et al 20..

$J(\theta)$ can be re-written as :

$$J(\theta) = -\log 4 + \mathbb{KL}\left[q(x) \,\|\, \frac{q(x) + p_\theta(x)}{2}\right] + \mathbb{KL}\left[p_\theta(x) \,\|\, \frac{q(x) + p_\theta(\dots}{2}\right.$$

Jensen - Shannon entropy

$$J(\theta) = -\log 4 + 2\, JSD\left[q(x) \,\|\, p_\theta(x)\right]$$

2.) If $G_\theta(z)$ and $D_\phi(x)$ have enough capacity, and at each step of the training algorithm $D_\phi(x)$ is allowed to reach its optimum, and $p_\theta(x)$ is updated to improve the following criterion:

$$\underbrace{\mathbb{E}_{x \sim q(x)}\left[\log D_\phi^*(x)\right] + \mathbb{E}_{z \sim p(z)}\left[\log\left(1 - D_\phi^*(G_\theta(z))\right)\right]}$$

then $p_\theta(x) \longrightarrow q(x)$.