Exercises for the course
Deep Learning 2
Summer Semester 2023

Machine Learning Group
Faculty IV – Electrical Engineering and Computer Science
Technische Universität Berlin

Exercise Sheet 4

Exercise 1: Global Optimality of the GAN objective (10 + 10 + 20 P)

In this exercise, we want to show that the global optimal solution for the minimax game

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{x}}(\boldsymbol{z})}[\log(1-D(G(\boldsymbol{z})))] \tag{1}$$

for training Generative Adversarial Networks is that the data distribution gained from sampling from p_g is equal to the real data distribution p_{data}.

(a) Therefore, we first consider the optimal discriminator D for any given generator G. Show that for fixed G, the optimal discriminator D is

$$D_G^*(\boldsymbol{x}) = \frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \tag{2}$$

Hint: For any $(a,b) \in \mathbb{R}^2 \backslash \{0,0\}$, $y \in [0,1]$, the function $f(y,a,b) = a\log(y) + b\log(1-y)$ achieves its maximum at $\frac{a}{a+b}$.

(b) Show that the maximum $C(G) = \max_D V(G,D)$ of the training criterion can be reformulated to:

$$C(G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}\left[\log \frac{p_{\text{data}}(\boldsymbol{x})}{P_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[\log \frac{p_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right] \tag{3}$$

(c) Show that the global minimum of $C(G)$ is $C^* = -\log(4)$ and that reaching it is equivalent to $p_g = p_{\text{data}}$.
Hint: Use the fact, that the the Jensen Shannon Divergence $JSD(P\|Q) = \frac{1}{2}(KL(P\|M) + KL(Q\|M))$ is always positive.

Exercise 2: Reformulating the loss function of diffusion models (20 P)

(a) Show that

$$L_{vlb} = \mathbb{E}_q\left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T} \mid \mathbf{x}_0)}\right]$$

can be reformulated to:

$$L_{vlb} = L_0 + L_1 + \ldots + L_{T-1} + L_T$$

where

$$L_0 = -\log p_\theta(x_0 \mid x_1)$$
$$L_{t-1} = D_{KL}(q(x_{t-1} \mid x_t, x_0) \| p_\theta(x_{t-1} \mid x_t))$$
$$L_T = D_{KL}(q(x_T \mid x_0) \| p(x_T))$$

with the help of the Markov assumption in Diffusion models.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.
Generate MNIST with diffusion model in PyTorch.