0 Instructions

Homework is due Tuesday, April 16, 2024 at 23:59pm Central Time. Please refer to `https://courses.grainger.illinois.edu/cs446/sp2024/homework/hw/index.html` for course policy on homeworks and submission instructions.

1 GAN: 7pts

1.1

When G is fixed, the objective function becomes

$$\begin{aligned}
&\max_D \mathbb{E}_{x\sim p_r(x)}[\log D(x)] + \mathbb{E}_{x\sim p_g(x)}[\log(1-D(x))] \\
\Leftrightarrow &\max_D \int p_r(x)\log D(x)dx + \int p_g(x)\log(1-D(x))dx \\
\Leftrightarrow &\max_D \int [p_r(x)\log D(x) + p_g(x)\log(1-D(x))]dx
\end{aligned}$$

For any given x, we can find a $D^*(x)$ that satisfies

$$D^*(x) = argmax_D\{p_r(x)\log D(x) + p_g(x)\log(1-D(x))\}$$

Take derivative to D, we get

$$\begin{aligned}
&\frac{p_r(x)}{D^*(x)} - \frac{p_g(x)}{1-D^*(x)} = 0 \\
\Longrightarrow &D^*(x) = \frac{p_r(x)}{p_g(x)+p_r(x)} \qquad (1)
\end{aligned}$$

1.2

When D is optimal, put equation (1) into objective function. We get

$$
\begin{aligned}
&\min_G \int (p_r \log \frac{p_r}{p_r + p_g} + p_g \log \frac{p_g}{p_r + p_g})dx \\
=&\min_G \int p_r \log p_r dx - \int p_r \log \frac{p_r + p_g}{2} dx - \int p_r \log 2dx \\
&+ \int p_g \log p_g dx - \int p_g \log \frac{p_r + p_g}{2} dx - \int p_g \log 2dx \\
\Leftrightarrow &\min_G \quad \frac{1}{2} D_{KL}(P_r || \frac{P_r + P_g}{2}) + \frac{1}{2} D_{KL}(P_g || \frac{P_r + P_g}{2})
\end{aligned}
$$

1.3

When D is perfect, the objective function will consistently equals to 0, which causes the gradients to 0, in other words, the problem of vanishing gradients.

2 Diffusion model : 11pts

2.1

$$\begin{aligned} ELBO &= \mathbb{E}_{q_\phi(x_1|x_0)} \log p_\theta(x_0|x_1) - D_{KL}(q_\phi(x_t|x_0)||p_{x_t}) \\ &\quad - \sum_{t=2}^{T} \mathbb{E}_{q_\phi(x_t|x_0)} D_{KL}(q_\phi(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t)) \\ &= \sum_{t=1}^{T} \frac{1}{2\sigma_q(t)^2} \frac{\beta_t \bar{\beta}_{t-1}}{(1-\bar{\beta}_t)^2} \mathbb{E}_{q_\phi(x_t|x_0)} ||\hat{x}_\theta(x_t) - x_0||^2 \end{aligned}$$

Where $\bar{\beta}_t = \Pi_{i=1}^t (1-\beta_i)$. The expectation is with respect to distribution $\mathcal{N}(\sqrt{\bar{\beta}_t} x_0, (1-\bar{\beta}_t)I)$.

2.2

No. The decoder takes input x_t from noise and then generates x_{t-1} from the previous output, and the process goes on. Because don't know the true distribution of each step, we can't directly estimate the density. But We can estimate it by sampling.

2.3

By using reparameterization trick, we can get

$$\begin{aligned} x_1 &= \sqrt{\beta_1}\epsilon_1 + \sqrt{1-\beta_1}x_0 \\ x_2 &= \sqrt{\beta_2}\epsilon_2 + \sqrt{1-\beta_2}x_1 \\ &\vdots \\ x_t &= \sqrt{\beta_t}\epsilon_t + \sqrt{1-\beta_t}x_{t-1} \end{aligned}$$

Where $\epsilon_i \sim \mathcal{N}(0, I)$.

Put all x_i (i = 1, 2, ..., t-1) into the x_t equation, we get

$$\begin{aligned} x_t &= \sqrt{1-\beta_t}\sqrt{1-\beta_{t-1}} \cdots \sqrt{1-\beta_1}x_0 \\ &+ \sum_{i=1}^{t-1} \sqrt{1-\beta_t}\sqrt{1-\beta_{t-1}} \cdots \sqrt{1-\beta_{i+1}}\sqrt{\beta_i}\epsilon_i + \sqrt{\beta_t}\epsilon_t \end{aligned}$$

x_t is from a Gaussian distribution with mean $\sqrt{1-\beta_t}\sqrt{1-\beta_{t-1}} \cdots \sqrt{1-\beta_1}x_0$ and variance

$$\bar{\epsilon}_t = (1-\beta_t)(1-\beta_{t-1})\cdots(1-\beta_2)\beta_1 + (1-\beta_t)(1-\beta_{t-1})\cdots(1-\beta_3)\beta_2 + \cdots + \beta_t$$

We can also get variances $\bar{\epsilon}_{t-1}$, $\bar{\epsilon}_{t-2}$, ..., $\bar{\epsilon}_1$ by the same way. We can simplify the equation by using the following equation:

$$\begin{aligned} \bar{\epsilon}_t &= (1-\beta_t)\epsilon_{t-1} + \beta_t \\ \Leftrightarrow \bar{\epsilon}_t - 1 &= (1-\beta_t)(\epsilon_{t-1} - 1) \\ \Rightarrow \bar{\epsilon}_t - 1 &= -\Pi_{i=1}^t (1-\beta_i) \\ \Rightarrow \bar{\epsilon}_t &= 1 - \Pi_{i=1}^t (1-\beta_i) \end{aligned}$$

So, $q(x_t|x_0) = \mathcal{N}(x_t; \Pi_{i=1}^t \sqrt{1-\beta_i}x_0, 1 - \Pi_{i=1}^t(1-\beta_i))$

2.4

We followed the method from Stanley H. Chan's "Tutorial on Diffusion Models for Imaging and Vision".

$$\begin{aligned} q_\phi(x_{t-1}|x_t, x_0) &= \frac{q_\phi(x_{t-1}, x_t|x_0)}{q_\phi(x_t|x_0)} \\ &= \frac{q_\phi(x_t|x_{t-1}, x_0)q_\phi(x_{t-1}|x_0)}{q_\phi(x_t|x_0)} \\ &= \frac{\mathcal{N}(x_t|\sqrt{1-\beta_t}x_{t-1}, \beta_t I)\mathcal{N}(x_{t-1}|\sqrt{\bar{\beta}_{t-1}}, (1-\bar{\beta}_{t-1})I)}{N(x_t|\sqrt{\bar{\beta}_t}x_0, (1-\bar{\beta}_t)I)} \end{aligned} \tag{2}$$

After calculation of above Gaussian product, we get

$$q_\phi(x_{t-1}|x_t, x_0) \sim exp\frac{x_t - \sqrt{1-\beta_t}x_{t-1}}{2\beta_t} + \frac{x_{t-1} - \sqrt{\bar{\beta}_{t-1}x_0}}{2(1-\bar{\beta}_{t-1})} - \frac{x_t - \sqrt{\bar{\beta}_t}x_0}{2(1-\bar{\beta}_t)}$$

Then, we do some variable substitution,

$$\begin{aligned} x &= x_t, \qquad a = \sqrt{1-\beta_t} \\ y &= x_{t-1}, \qquad b = \sqrt{\bar{\beta}_{t-1}} \\ z &= x_0, \qquad c = \sqrt{\bar{\beta}_t} \end{aligned}$$

Then, we get

$$f(y) = \frac{x-\sqrt{a}y}{2(1-a)} + \frac{y-\sqrt{b}z}{2(1-b)} - \frac{x-\sqrt{c}z}{2(1-c)}$$

The mean of $q_\phi(x_{t-1}|x_t, x_0)$ is where the derivative of $f(y)$ equals to 0,

$$f'(y) = \frac{1-ab}{(1-a)(1-b)}y - (\frac{\sqrt{a}}{1-a}x + \frac{\sqrt{b}}{1-b}z) = 0$$

which yeilds

$$y = \frac{(1-b)\sqrt{a}}{1-ab}x + \frac{(1-a)\sqrt{b}}{1-ab}z$$

So, the mean of $q_\phi(x_{t-1}|x_t, x_0)$ is,

$$\mu_q(x_t, x_0) = \frac{(1-\bar{\beta}_{t-1})\sqrt{1-\beta_t}}{1-\bar{\beta}_t}x_t + \frac{\beta_t\sqrt{\bar{\beta}_{t-1}}}{1-\bar{\beta}_t}x_0$$

where $\bar{\beta}_t = \Pi_{i=1}^t(1-\beta_i)$.

2.5

$$\begin{aligned}
s_\theta(x, \delta|x_{known}) &= \nabla_x \log p_\theta(x, \delta|x_{known}) \\
&= \nabla_x \log \frac{p(x_{known}|x)p_\theta(x, \delta)}{p(x_{known})} \\
&= \nabla_x[\log p(x_{known}|x) + \log p_\theta(x, \delta) - \log p(x_{known})] \\
&= \nabla_x[-||(x - x_{known}) \odot M||^2] + s_\theta(x, \delta) \\
&= -2((x - x_{known}) \odot M) + s_\theta(x, \delta)
\end{aligned}$$

3 Unsupervised learning / contrastive learining: 4 pts

3.1

True.

3.2

False. MAE uses different mask-out rates during training.

3.3

True.

3.4

False. When using CLIP to clssify images, we can insert labels into appropriate sentences for feature extraction, compared with the features of images. If the similarity is high, it can be considered that the labels contained in the sentence are the labels of the image.

4 Coding: GAN, 10pts

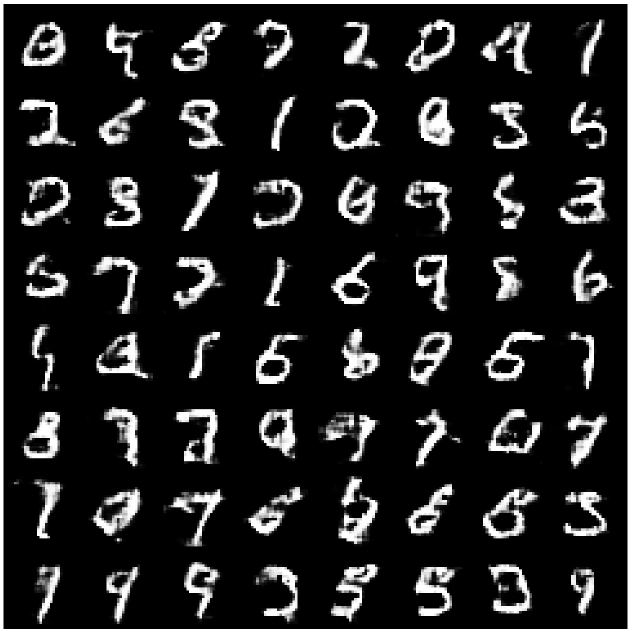

Figure 1: Generated images after 90 epochs

5 Coding: Diffusion model, 10pts

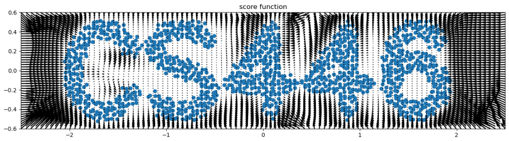

Figure 2: Score function

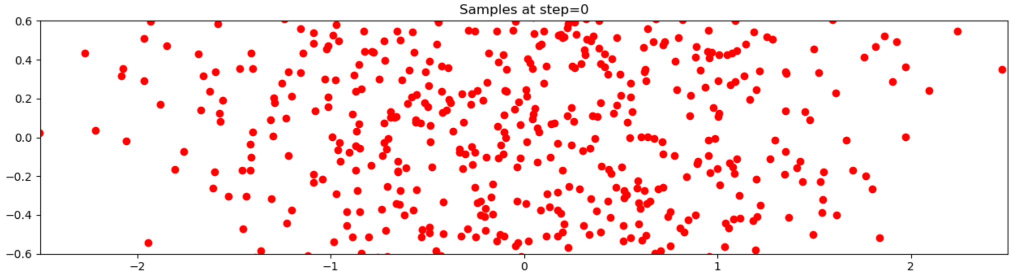

Figure 3: Step 0

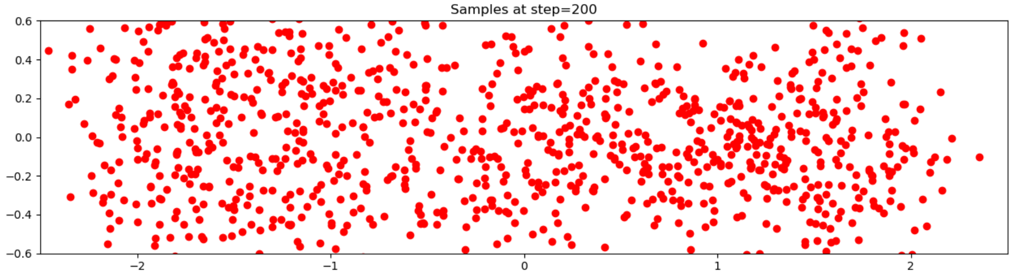

Figure 4: Step 200

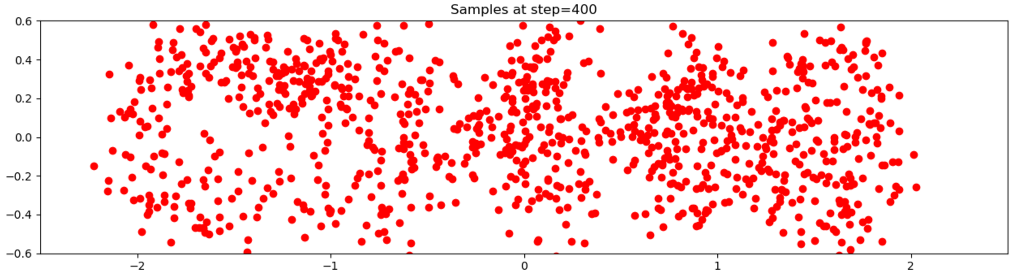

Figure 5: Step 400

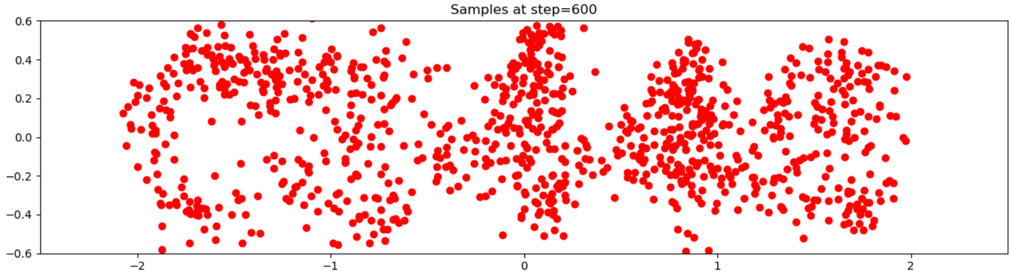

Figure 6: Step 600

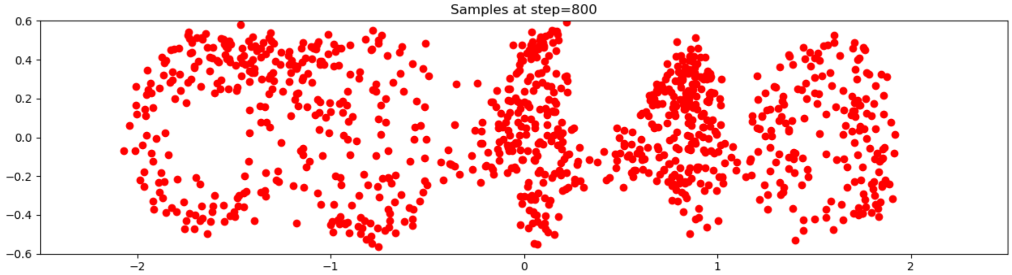

Figure 7: Step 800

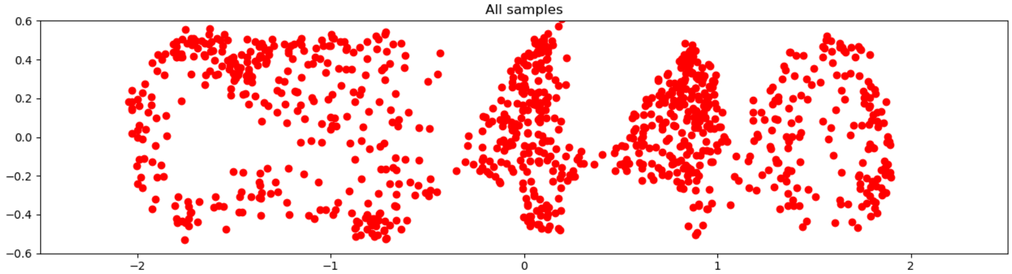

Figure 8: Final

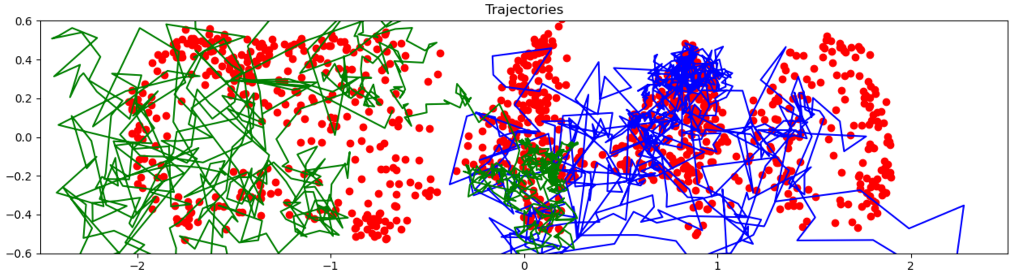

Figure 9: Trajectory