Homework 5

Graded

Student

Jinzhi Shen

Total Points

64.5 / 82 pts

Question 1

VAE **22** / 35 pts

1.1 **1.a** **0** / 3.5 pts

− **0 pts** Correct

✔ − **3.5 pts** Incorrect

− **1 pt** Incorrect notation

Use y_j in the answer. Also missing product over G.

1.2 **1.b** **3.5** / 3.5 pts

✔ − **0 pts** Correct

− **3.5 pts** Incorrect

− **1 pt** Incorrect dimension

− **2.5 pts** Incorrect explanation

1.3 **1.c** **3.5** / 3.5 pts

✔ − **0 pts** Correct

− **3.5 pts** Incorrect

− **2.5 pts** Major mistake

− **1 pt** Minor Mistake

1.4 **1.d** **3.5** / 3.5 pts

✔ − **0 pts** Correct

− **1.75 pts** One property incorrect

− **3.5 pts** Both properties incorrect

1.5 **1.e** **3.5** / 3.5 pts

✔ − **0 pts** Correct

− **3.5 pts** Incorrect

− **1.75 pts** Explanation missing

1.6 **1.f** **3.5** / 3.5 pts

✔ − **0 pts** Correct

− **2.5 pts** Explanation not provided

− **3.5 pts** Incorrect

1.7 **1.g** **3.5** / 3.5 pts

- ✔ **– 0 pts** Correct
- **– 3.5 pts** Missing
- **– 2.5 pts** Major Mistake
- **– 1 pt** Minor Mistake

1.8 **1.h** **1** / 3.5 pts

- **– 0 pts** Correct
- **– 3.5 pts** Missing
- ✔ **– 2.5 pts** Major Mistake
- **– 1 pt** Minor Mistake

1.9 **1.i** **0** / 3.5 pts

- **– 0 pts** Correct
- ✔ **– 3.5 pts** Missing
- **– 2.5 pts** Major Mistake
- **– 1 pt** Minor Mistake

1.10 **1.j** **0** / 3.5 pts

- **– 0 pts** Correct
- ✔ **– 3.5 pts** Incorrect

Question 2

VAE Coding **8** / 9 pts

2.1 **2.e.i** **2** / 3 pts

– **0 pts** Correct

– **3 pts** Figure not provided

– **1 pt** Low sample count (<4)

✔ – **1 pt** Samples do not resemble digits

– **0.5 pts** Either logits or bernoulli samples are not provided

2.2 **2.e.ii** **3** / 3 pts

✔ – **0 pts** Correct

– **3 pts** Figure not provided or not correct

2.3 **2.e.iii** **3** / 3 pts

✔ – **0 pts** Correct

– **3 pts** Figure not provided or incorrect

– **1.5 pts** Interpolation is partially demonstrated (e.g., only providing interpolation at alpha=0.5)

Question 3

GAN **10.5** / 14 pts

3.1 **3.a** **3.5** / 3.5 pts

- ✔ **– 0 pts** Correct
- **– 3.5 pts** Incorrect

3.2 **3.b** **0** / 3.5 pts

- **– 0 pts** Correct
- ✔ **– 3.5 pts** Incorrect
- **– 1 pt** Minor mistake

3.3 **3.c** **3.5** / 3.5 pts

- ✔ **– 0 pts** Correct
- **– 3.5 pts** Missing/Wrong
- **– 2.5 pts** Major Mistake
- **– 1 pt** Minor Mistake

3.4 **3.d** **3.5** / 3.5 pts

- ✔ **– 0 pts** Correct
- **– 3.5 pts** Missing/Wrong
- **– 2.5 pts** Major Mistake
- **– 1 pt** Minor Mistake

Question 4

Diffusion Models **24** / 24 pts

4.1 **4.b** **12** / 12 pts

- ✔ **– 0 pts** Correct
- **– 12 pts** No code and no figure
- **– 6 pts** Code provided, gradient vector field does not resemble data

4.2 **4.c** **12** / 12 pts

- ✔ **– 0 pts** Correct
- **– 12 pts** No code and no figure
- **– 6 pts** Code Provided, samples do not match distribution
- **– 2 pts** Code Provided, samples roughly match distribution (e.g., samples are grouped into a subset or are very thinly distributed)

Questions assigned to the following page: 1.1, 1.2, 1.3, 1.4, 1.5, and 1.6

Homework 5

Question 1

(a)

Denote the probability of the Bernoulli distribution associated with z as $\theta(z)$, then the explicit form for $p_\theta(x|z)$ is

$$\hat{y}_j = \theta(z)^{x^j}(1-\theta(z))^{1-x^j}$$

(b)

The output dimension of the encoder is 2 since the dimension of the latent space is 2.

(c)

Using Jensen's inequality to obtain a bound on the log-likelihood:

$$\begin{aligned}
log\ p_\theta(x) &= log \int p_\theta(x,z)dz \\
&= log \int q_\phi(z|x)\frac{p_\theta(x,z)}{q_\phi(z|x)}dz \\
&\geq \int q_\phi(z|x) log \frac{p_\theta(x,z)}{q_\phi(z|x)}dz \quad \textit{(Jensen's inequality)} \\
&= \mathcal{L}(p_\theta, q_\phi) \quad (ELBO)
\end{aligned} \tag{1}$$

Dividing the bound into two parts, one of which is the Kullback-Leibler divergence $KL(q_\phi(z|x), p(z))$:

$$\begin{aligned}
\mathcal{L}(p_\theta, q_\phi) &= \int q_\phi(z|x) log \frac{p_\theta(x,z)}{q_\phi(z|x)}dz \\
&= \int q_\phi(z|x) log \frac{p_\theta(x|z)p(z)}{q_\phi(z|x)}dz \\
&= \int q_\phi(z|x) log \frac{p(z)}{q_\phi(z|x)}dz + \int q_\phi(z|x) log\ p_\theta(x|z)dz \\
&= -KL(q_\phi(z|x), p(z)) + \int q_\phi(z|x) log\ p_\theta(x|z)dz
\end{aligned} \tag{2}$$

(d)

1. KL-divergence is non-negative
2. KL-divergence is not symmetric, which means $D_{KL}(P||Q) \neq D_{KL}(Q||P)$

(e)

They are not the same. Eq.(2) is more computational efficient since there is no need to separately sample from distribution $q_\phi(z|x)$ to compute the KL-divergence.

(f)

It is not a good idea to choose $q_\phi(z|x) := \mathcal{N}(0, \mathcal{I})$ because in that way z doesn't contain information about x and is not a good representation for x.

Questions assigned to the following page: 2.1, 2.2, 1.7, 1.8, and 1.10

(g)

The value of KL-divergence $KL(q_\phi(z|x), q_\phi(z|x))$ is 0 because:

$$KL(q_\phi(z|x), q_\phi(z|x)) = -\sum_z q_\phi(z|x) log \frac{q_\phi(z|x)}{q_\phi(z|x)} = -\sum_z q_\phi(z|x) log(1) = 0$$

(h)

$$\begin{aligned} KL(q_\phi(z|x), p(z)) &= -\sum_z q_\phi(z|x) log \frac{p(z)}{q_\phi(z|x)} \\ &= -\sum_z \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(z-\mu_\phi)^2}{2\sigma^2}\right) log\left[exp\left(-\frac{(z-\mu_p)^2}{2\sigma^2} + \frac{(z-\mu_\phi)^2}{2\sigma^2}\right)\right] \\ &= -\sum_z \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(z-\mu_\phi)^2}{2\sigma^2}\right) \frac{(z-\mu_\phi)^2 - (z-\mu_p)^2}{2\sigma^2} \end{aligned} \tag{3}$$

(j)

(3) $p_\theta(x|z)$

Question 2

(e)

(i)

The corresponding screenshot is shown in Figure 1.

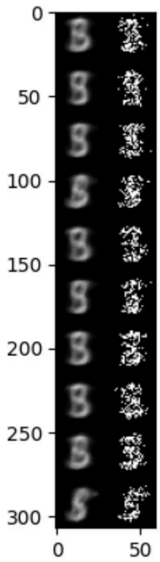

Figure 1: Question 2(e)i plot.

(ii)

The corresponding screenshot is shown in Figure 2.

Questions assigned to the following page: 3.1, 3.2, 2.2, 2.3, and 3.3

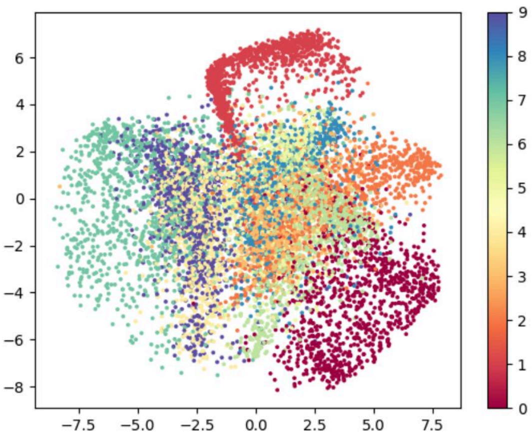

Figure 2: Question 2(e)ii plot.

(iii)

The corresponding screenshot is shown in Figure 3.

Question 3

(a)

The cost function is:

$$\max_{\theta} \min_{w} - \sum_{x} log\ D_w(x) - \sum_{z} log\ (1 - D_w(G_\theta(z)))$$

(b)

Assuming arbitrary capacity:

$$\begin{aligned} \min_{D} : \quad & - \int_x p_{data}(x) log\ D(x) dx - \int_z p_z(z) log\ (1 - D(G_\theta(z))) dz \\ & = - \int_x p_{data}(x) log\ D(x) + p_G(x) log\ (1 - D(x)) dx \end{aligned} \tag{4}$$

(c)

Euler-Lagrange formalism:

$$S(D) = \int_x L(x, D, \dot{D}) dx$$

From

$$\frac{\partial L(x, D, \dot{D})}{\partial D} - \frac{d}{dx} \frac{\partial L(x, D, \dot{D})}{\partial \dot{D}} = 0$$

and $\frac{d}{dx} \frac{\partial L(x,D,\dot{D})}{\partial \dot{D}}$ can be removed,
we have

$$\frac{\partial L(x, D, \dot{D})}{\partial D} = - \frac{p_{data}}{D} + \frac{p_G}{1 - D} = 0 \implies D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_G(x)}$$

Questions assigned to the following page: 4.1, 4.2, 3.3, and 3.4

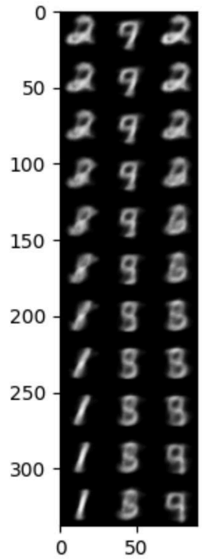

Figure 3: Question 2(e)iii plot.

(d)

Assume arbitrary capacity and an optimal discriminator $D^*(x)$,

$$\begin{aligned}
&-\int_x p_{data} log\ D^*(x) + p_G(x) log\ (1 - D^*(x)) dx \\
&= -\int_x p_{data}(x) log \frac{p_{data}(x)}{p_{data}(x) + p_G(x)} + p_G(x) log \frac{p_G x}{p_{data}(x) + p_G(x)} dx \\
&= -2JSD(p_{data}, p_G) + log(4)
\end{aligned} \tag{5}$$

Therefore, the optimal generator $G^*(x)$ generates the distribution $p_G^* = p_{data}$

Question 4

(b)

The corresponding screenshot is shown in Figure 4.

(c)

The corresponding screenshot is shown in Figure 5.

Question assigned to the following page: 4.2

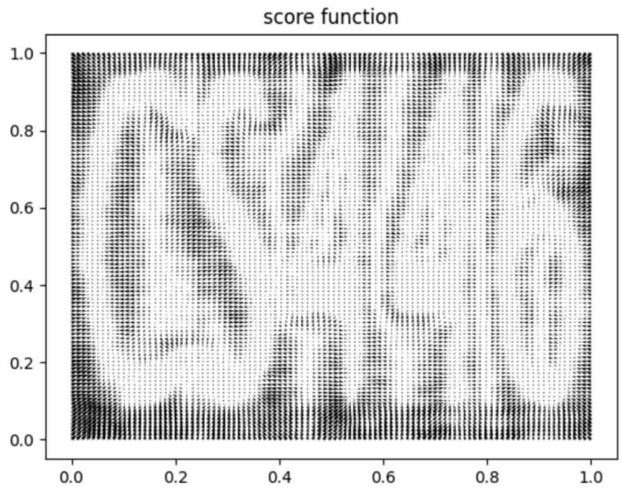

Figure 4: Question 4(b) screenshot.

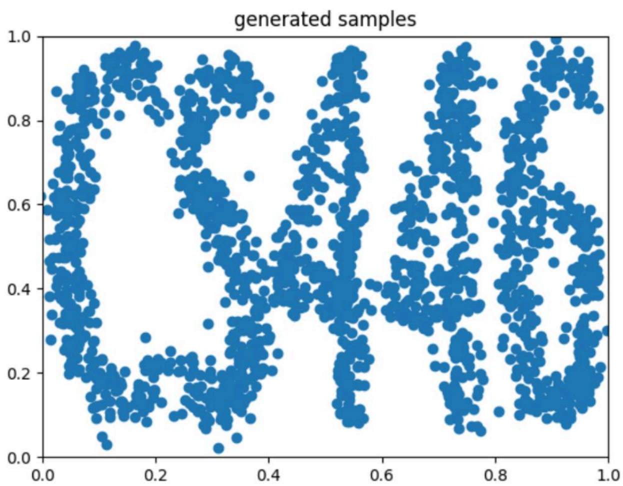

Figure 5: Question 4(c) screenshot.