MFDNN HW10

MinGyu Shin

May 27, 2024

Problem 1 : *Log-derivative trick for VAE.*

$$\begin{aligned}
\frac{\partial}{\partial \phi_i}\mathbb{E}_{Z\sim q_\phi(z)}\left[\log\left(\frac{h(Z)}{q_\phi(Z)}\right)\right] &= \frac{\partial}{\partial \phi_i}\int \log\left(\frac{h(z)}{q_\phi(z)}\right) q_\phi(z)dz \\
&= \int \frac{\partial}{\partial \phi_i}\left(\log\left(\frac{h(z)}{q_\phi(z)}\right) q_\phi(z)\right) dz \\
&= \int -\frac{1}{q_{\phi(z)}}\left(\frac{\partial}{\partial \phi_i} q_{\phi(z)}\right) q_{\phi(z)} + \log\left(\frac{h(z)}{q_\phi(z)}\right)\left(\frac{\partial}{\partial \phi_i} q_{\phi(z)}\right) dz \\
&= \int \log\left(\frac{h(z)}{q_\phi(z)}\right)\left(\frac{1}{q_\phi(z)}\frac{\partial}{\partial \phi_i} q_\phi(z)\right) q_\phi(z)dz \\
&= \mathbb{E}_{Z\sim q_\phi(Z)}\left[\left(\frac{\partial}{\partial \phi_i}\log(q_\phi(Z))\right)\log\left(\frac{h(Z)}{q_\phi(Z)}\right)\right].
\end{aligned}$$

$$\Rightarrow \nabla_\phi \mathbb{E}_{Z\sim q_\phi(z)}\left[\log\left(\frac{h(Z)}{q_\phi(Z)}\right)\right] = \mathbb{E}_{Z\sim q_\phi(Z)}\left[(\nabla_\phi \log(q_\phi(Z))\log\left(\frac{h(Z)}{q_\phi(Z)}\right)\right].$$

Problem 2 : *Projected gradient method.*

$$\Pi_C(y) = \underset{x\in C}{\operatorname{argmin}}\, \|x-y\|^2 = \underset{x\in C}{\operatorname{argmin}}\, (x_1-y_1)^2 + (x_2-y_2)^2.$$

Since a is an only choice to be in C, x_1 should be a. If $0 \le y_2 \le 1$, then $x_2 = y_2$ is optimal. If $y_2 < 0$ or $y_2 > 1$, the closer one of $0, 1$ is optimal. Thus

$$\Pi_C(y) = \begin{bmatrix} a \\ \min\{\max\{y_2, 0\}, 1\}. \end{bmatrix}$$

Problem 3 : *Image inpainting with flow models.*

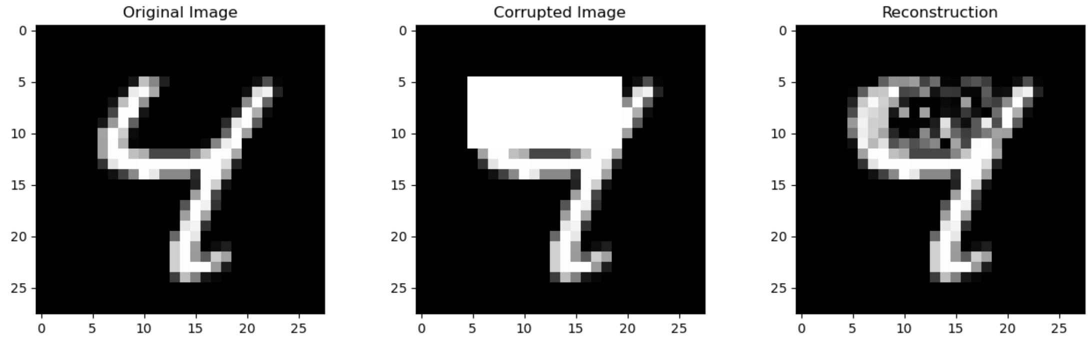


Figure 1: The original, corrupted, and inpainted MNIST images.

Problem 4 : *Ingridents of Glow.*

(a) Note that

$$|P| = 1 \quad |L| = 1 \quad |U + \mathrm{diag}(s)| = \prod_{i=1}^{C} |s_i|.$$

Thus,

$$\log\left|\frac{\partial f_1}{\partial x}\right| = \log|A| = \log|P||L||U + \mathrm{diag}(s)| = \log\prod_{i=1}^{C}|s_i| = \sum_{i=1}^{C}\log s_i.$$

(b) For consistent reshape operation, there are only difference in order of rows and columns. It can be represented by finite multiplication of left and right multiplication of permutation matrices, which preserves the determinant.

(c) Let reshape operation as for each location of layer, ordering layerwise first and after considering next location. Then,

$$\left|\frac{\partial f_2(X|P,L,U,s)}{\partial X}\right| = |\mathrm{diag}(A)_{mn}|$$

where the RHS is block diagonal matrix with mn times of A. Thus by block determinant cal,

$$\log\left|\frac{\partial f_2(X|P,L,U,s)}{\partial X}\right| = \log(\prod_{i=1}^{C}|s_i|)^{mn} = mn\sum_{i=1}^{C}\log|s_i|.$$

Problem 5 : *Gambler's ruin.*

With prob 18/37, the estimator is inaccurate. But with importance sampling, the estimator is more accurate with prob 0.55.

```
Without IS,  0.0
With IS,  2.6414300427861382e-06
```

Figure 2: Gambler's ruin.

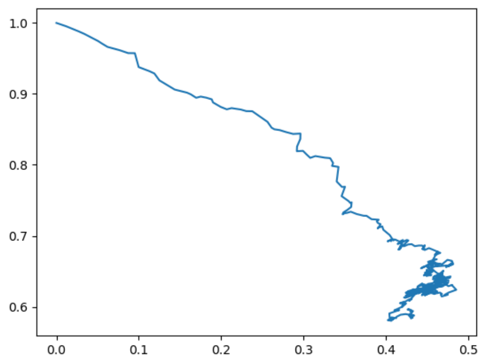
Problem 6 : *Practice of log-derivatives trick and reparameterization trick.*

For each trick, 500 iterations, learning rate $= 1e-2$ and 16 batch size are used. Reparameterization trick seems to be much stable.

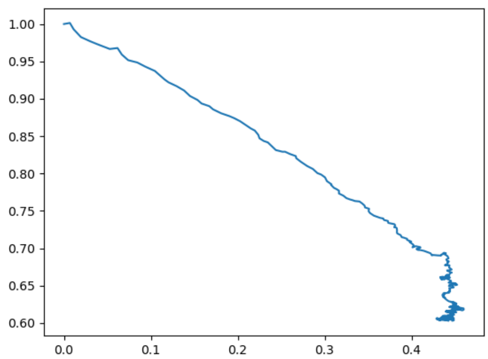
```
Log-derivative trick
mu : 0.4298, sigma : 0.5953

Reparameterization trick
mu : 0.4510, sigma : 0.6067
```

Figure 3: Estimator for mu and sigma.



(a) Log-derivatives trick.



(b) Reparameterization trick.