MFDNN HW9

MinGyu Shin

May 17, 2024

Problem 1 : *Anomaly detection via AE.*

The training loss, the threshold & error rate and error samples are below.

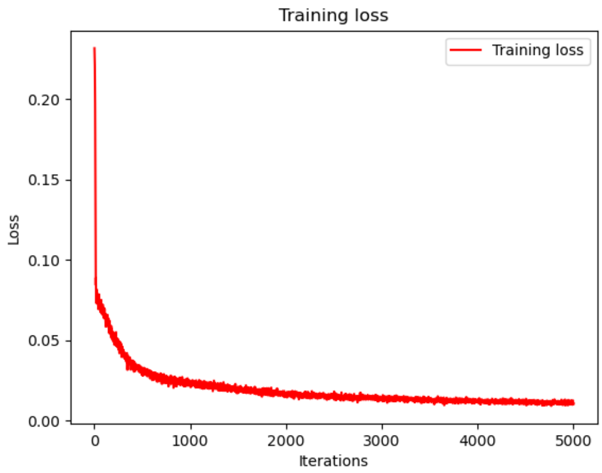


Figure 1: Training loss

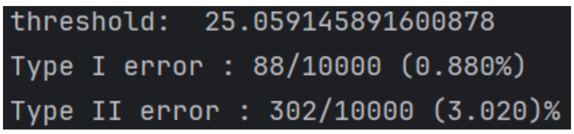
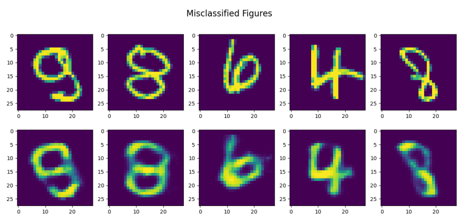
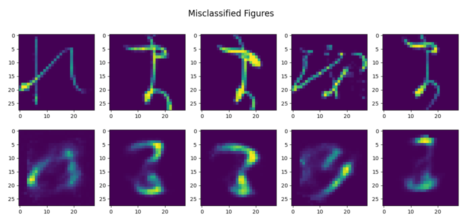


Figure 2: Training loss



(a) Type 1 Error Sample



(b) Type 2 Error Sample

Problem 2 : *1D flow to Gaussian.*

Training and test loss, p_X, p_Z are below.

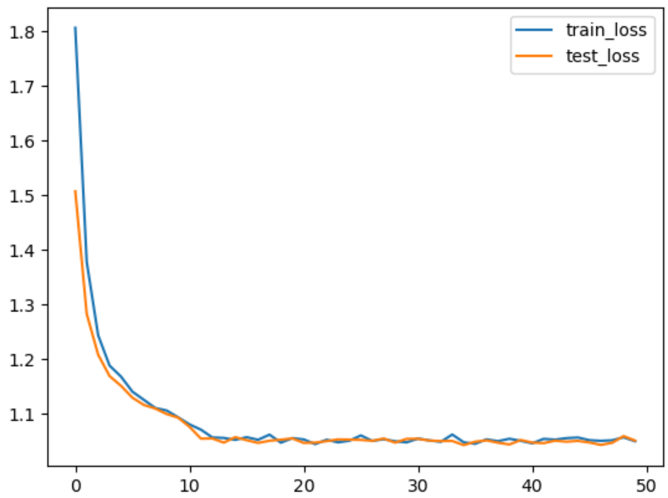
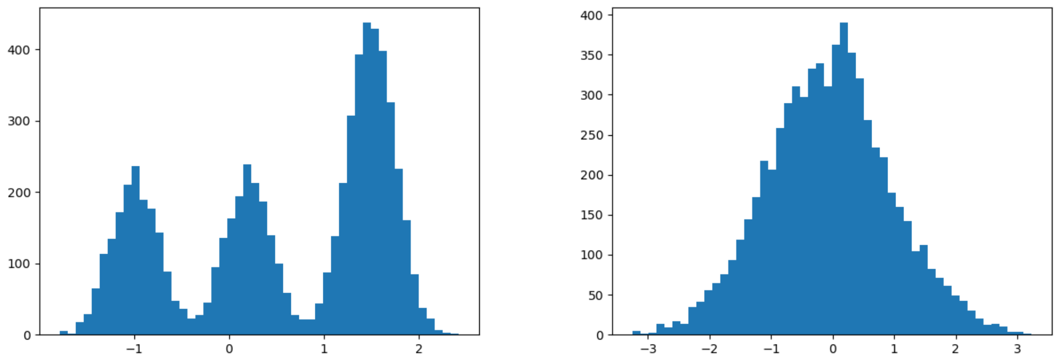


Figure 4: Training and test loss.



(a) p_X (b) p_Z

Problem 3 : *Affine coupling layer with permutations.*

Let σ such that for ordered $\Omega, \Omega^C, P_\sigma \begin{bmatrix} 1 \\ \vdots \\ n \end{bmatrix} = \begin{bmatrix} \Omega \\ \Omega^C \end{bmatrix} = U$. Then

$$(P_\sigma \frac{\partial z}{\partial x} P_\sigma^{-1})_{ij} = (P_\sigma)_i \frac{\partial z}{\partial x} (P_\sigma)_j^\intercal = e_{\sigma(i)} \frac{\partial z}{\partial x} e_{\sigma(j)} = \frac{\partial z_{\sigma(i)}}{\partial x_{\sigma(j)}} = \frac{\partial z_{U(i)}}{\partial x_{U(j)}}.$$

Thus,

$$P_\sigma \frac{\partial z}{\partial x} P_\sigma^{-1} = \begin{bmatrix} I & 0 \\ \frac{z_{\Omega^C}}{x_\Omega} & \frac{z_{\Omega^C}}{x_{\Omega^C}} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \frac{z_{\Omega^C}}{x_\Omega} & \operatorname{diag}(e^{s_\theta(x_\Omega)}) \end{bmatrix},$$

$$\Rightarrow \log|\frac{\partial z}{\partial x}| = \log(|P_{\sigma^{-1}} \begin{bmatrix} I & 0 \\ \frac{z_{\Omega^C}}{x_\Omega} & \operatorname{diag}(e^{s_\theta(x_\Omega)}) \end{bmatrix} P_\sigma|) = \log|\prod_{i=1}^{n-|\Omega|} e^{s_\theta(x_\Omega)_i}| = \mathbf{1}_{n-|\Omega|}^\intercal s_\theta(x_\Omega),$$

since $|P_\sigma| = |P_{\sigma^{-1}}| = 1$.

Problem 4 : D_{KL} *of continuous random variables.*

(a) Since

$$-D_{\mathrm{KL}}(X\|Y) = \int_{\mathbb{R}^d} f(x)\log(\frac{g(x)}{f(x)})dx = \mathop{\mathbb{E}}_X[\log(\frac{g(x)}{f(x)})] \leq \log(\mathop{\mathbb{E}}_X[\frac{g(x)}{f(x)}]) = 0,$$

$$D_{\mathrm{KL}}(X\|Y) \geq 0.$$

(b)

$$\begin{aligned} D_{\mathrm{KL}}(X\|Y) &= \int_{X_d}\cdots\int_{X_1} f_X(x_1,\ldots,x_d)\log(\frac{f_X(x_1,\ldots,x_d)}{g_Y(x_1,\ldots,x_d)})dx_1\ldots dx_d \\ &= \int_{X_d}\cdots\int_{X_1} f_{X_1}(x_1)\ldots f_{X_d}(x_d)(\log(\frac{f_{X_1}(x_1)}{g_{Y_1}(x_1)}) + \cdots + \log(\frac{f_{X_d}(x_d)}{g_{Y_d}(x_d)}))dx_1\ldots dx_d \\ &= \int_{X_1} f_{X_1}(x_1)\log(\frac{f_{X_1}(x_1)}{g_{Y_1}(x_1)})dx_1 + \cdots + \int_{X_d} f_{X_d}(x_d)\log(\frac{f_{X_d}(x_d)}{g_{Y_d}(x_d)})dx_d \\ &= D_{\mathrm{KL}}(X_1\|Y_1) + \cdots + D_{\mathrm{KL}}(X_d\|Y_d), \end{aligned}$$

since $X_1,\ldots,X_d$ and $Y_1,\ldots,Y_d$ are independent respectively.

Problem 5 : D_{KL} *of Gaussian random variables.*

$$\begin{aligned} D_{\mathrm{KL}}(X\|Y) &= \int_{\mathbb{R}^d} f(x)(\log(f(x) - \log(g(x)))dx \\ &= \int_{\mathbb{R}^d} f(x)[-\frac{1}{2}\log((2\pi)^d|\Sigma_0|) - \frac{1}{2}(x-\mu_0)^\intercal\Sigma_0^{-1}(x-\mu_0) \\ &\quad + \frac{1}{2}\log((2\pi)^d|\Sigma_1|) + \frac{1}{2}(x-\mu_1)^\intercal\Sigma_1^{-1}(x-\mu_1)]dx \\ &= \frac{1}{2}\underbrace{\mathbb{E}[\log(|\Sigma_1| - |\Sigma_0|)]}_{(1)} - \frac{1}{2}\underbrace{\mathbb{E}[(x-\mu_0)^\intercal\Sigma_0^{-1}(x-\mu_0)]}_{(2)} + \frac{1}{2}\underbrace{\mathbb{E}[(x-\mu_1)^\intercal\Sigma_1^{-1}(x-\mu_1)]}_{(3)}. \end{aligned}$$

$$(1) = \log(\frac{|\Sigma_1|}{|\Sigma_0|}),$$

$$\begin{aligned}
(2) &= \mathbb{E}[\mathrm{Tr}((x-\mu_0)^\intercal \Sigma_0^{-1}(x-\mu_0))] \\
&= \mathbb{E}[\mathrm{Tr}(\Sigma_0^{-1}(x-\mu_0)(x-\mu_0)^\intercal)] \\
&= \mathrm{Tr}(\mathbb{E}[\Sigma_0^{-1}(x-\mu_0)(x-\mu_0)^\intercal]) \\
&= \mathrm{Tr}(\Sigma_0^{-1}\Sigma_0) = \mathrm{Tr}(I_d) = d,
\end{aligned}$$

$$\begin{aligned}
(3) &= \mathbb{E}[(x-\mu_1)^\intercal \Sigma_1^{-1}(x-\mu_1)] \\
&= \mathbb{E}[((x-\mu_0)-(\mu_1-\mu_0))^\intercal \Sigma_1^{-1}((x-\mu_0)-(\mu_1-\mu_0))] \\
&= \mathbb{E}[(x-\mu_0)^\intercal \Sigma_1^{-1}(x-\mu_0)] - 2\mathbb{E}[(\mu_1-\mu_0)^\intercal \Sigma_1^{-1}(x-\mu_0)] + \mathbb{E}[(\mu_1-\mu_0)^\intercal \Sigma_1^{-1}(\mu_1-\mu_0)] \\
&\overset{*}{=} \mathrm{Tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1-\mu_0)^\intercal \Sigma_1^{-1}(\mu_1-\mu_0).
\end{aligned}$$

Thus,

$$D_{\mathrm{KL}}(X\|Y) = \frac{1}{2}(\log(\frac{|\Sigma_1|}{|\Sigma_0|}) - d + \mathrm{Tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1-\mu_0)^\intercal \Sigma_1^{-1}(\mu_1-\mu_0)).$$

Problem 6 : *When maximizing a lower bound is tight.*

Let $\theta_0 \in \operatorname{argmax} f$. Since $\exists\, \phi_0 \in \Phi$ such that $h(\theta_0, \phi_0) = 0$,

$$\begin{aligned}
g(\theta_0, \phi_0) = f(\theta_0) &= \max_{\theta\in\Theta} f(\theta) \\
&= \max_{\theta\in\Theta} g(\theta,\phi) + h(\theta,\phi) \quad \forall\phi\in\Phi \\
&= \max_{\theta\in\Theta} g(\theta,\phi_0),
\end{aligned}$$

that is, $(\theta_0, \phi_0) \in \operatorname{argmax} g$.

Conversely, let $(\theta, \phi_0) \in \operatorname*{argmax}_{\phi\in\Phi} g(\theta,\phi)$ for arbitrary $\theta \in \Theta$. For $\phi_1 \in \Phi$ such that $h(\theta, \phi_1) = 0$,

$$g(\theta,\phi_0) \le g(\theta,\phi_0) + h(\theta,\phi_0) = g(\theta,\phi_1) + h(\theta,\phi_1) = g(\theta,\phi_1),$$

that is, $h(\theta, \phi_0) = 0$. Thus, for $(\theta_0, \phi_0) \in \operatorname{argmax} g(\theta, \phi)$,

$$\begin{aligned}
f(\theta_0) = g(\theta_0,\phi_0) &= \max_{\theta\in\Theta,\phi\in\Phi} g(\theta,\phi) \\
&= \max_{\theta\in\Theta} (\max_{\phi\in\Phi} g(\theta,\phi)) \\
&= \max_{\theta\in\Theta} (\max_{\phi\in\Phi} g(\theta,\phi) + h(\theta,\phi)) \\
&= \max_{\theta\in\Theta} f(\theta),
\end{aligned}$$

that is, $\theta_0 \in \operatorname{argmax} f$.