NTU CSIE FAI 2024 Homework 4

1. Hand-written Part

1.

$$\varphi(s) = s \cdot \theta(s)$$

$$\varphi'(s) = s' \theta(s) + s \cdot \theta'(s) = \frac{1}{1+e^{-s}} + \frac{se^{-s}}{(1+e^{-s})^2}$$

2.

(A)

$$V_1 = \begin{bmatrix} 0, & 1, & 0.5 \\ 0, & 0, & 0.5 \\ 1, & 0, & 0 \end{bmatrix} \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix} = \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix}$$

$$V_2 = \begin{bmatrix} 0, & 1, & 0.5 \\ 0, & 0, & 0.5 \\ 1, & 0, & 0 \end{bmatrix} \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix} = \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix}$$

$$V_3 = \begin{bmatrix} 0, & 1, & 0.5 \\ 0, & 0, & 0.5 \\ 1, & 0, & 0 \end{bmatrix} \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix} = \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix}$$

$$V_4 = \begin{bmatrix} 0, & 1, & 0.5 \\ 0, & 0, & 0.5 \\ 1, & 0, & 0 \end{bmatrix} \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix} = \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix}$$

$$V_5 = \begin{bmatrix} 0, & 1, & 0.5 \\ 0, & 0, & 0.5 \\ 1, & 0, & 0 \end{bmatrix} \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix} = \begin{bmatrix} \text{illegible} \\ \text{illegible} \\ \text{illegible} \end{bmatrix}$$

(B)

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0, & 1, & 0.5 \\ 0, & 0, & 0.5 \\ 1, & 0, & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} b+0.5c \\ 0.5c \\ a \end{bmatrix}$$

$\Rightarrow a = c,\ b = 0.5c,\ a+b+c = 1$

$$V^* = \begin{bmatrix} \frac{2}{5} \\ \frac{1}{5} \\ \frac{2}{5} \end{bmatrix}$$

3.

(A)

R_1 : $(1,2) \Rightarrow \mu_1$

$(3,4) \Rightarrow \mu_2$

$(7,0) \Rightarrow \mu_2 \quad (7-1)^2+(0-2)^2 > (7-3)^2+(0-4)^2$

$(10,2) \Rightarrow \mu_2 \quad (10-1)^2+(2-2)^2 > (10-3)^2+(2-4)^2$

$\mu_1 = (1,2)$, $\mu_2 = \left(\frac{3+7+10}{3}, \frac{4+0+2}{3}\right) = \left(\frac{20}{3}, 2\right)$

R_2 : $(1,2) \Rightarrow \mu_1$

$(3,4) \Rightarrow \mu_1 \quad (3-1)^2+(4-2)^2 < (3-\frac{20}{3})^2+(4-2)^2$

$(7,0) \Rightarrow \mu_2 \quad (7-1)^2+(0-2)^2 > (7-\frac{20}{3})^2+(0-2)^2$

$(10,2) \Rightarrow \mu_2 \quad (10-1)^2+(2-2)^2 > (10-\frac{20}{3})^2+(2-2)^2$

$\mu_1 = \left(\frac{1+3}{2}, \frac{2+4}{2}\right) = (2,3)$ $\mu_2 = \left(\frac{7+10}{2}, \frac{0+2}{2}\right) = \left(\frac{17}{2}, 1\right)$

R_3 : $(1,2) \Rightarrow \mu_1 \quad (1-2)^2+(2-3)^2 < (1-\frac{17}{2})^2+(2-1)^2$

$(3,4) \Rightarrow \mu_1 \quad (3-2)^2+(4-3)^2 < (3-\frac{17}{2})^2+(4-1)^2$

$(7,0) \Rightarrow \mu_2 \quad (7-2)^2+(0-3)^2 > (7-\frac{17}{2})^2+(0-1)^2$

$(10,2) \Rightarrow \mu_2 \quad (10-2)^2+(2-3)^2 > (10-\frac{17}{2})^2+(2-1)^2$

Converge !

(B)

R_1 : $(1,2) \Rightarrow \mu_1$

$(3,4) \Rightarrow \mu_1$ $\quad (3-1)^2+(4-2)^2 < (3-7)^2+(4-0)^2$

$(7,0) \Rightarrow \mu_2$

$(10,2) \Rightarrow \mu_2$ $\quad (10-1)^2+(2-2)^2 > (10-7)^2+(2-0)^2$

$\mu_1 = \left(\frac{1+3}{2}, \frac{2+4}{2}\right) = (2,3)$ $\quad \mu_2 = \left(\frac{7+10}{2}, \frac{0+2}{2}\right) = \left(\frac{17}{2}, 1\right)$

R_2 : Same as (A) R_3 $\Rightarrow$ Converge!

Result same as (A).

(C)

if centroids $\{\mu_1, \mu_2\} = \{(2,3), (\frac{22}{3}, \frac{8}{3})\}$

$(1,2), (3,4) \Rightarrow \mu_1$

$(5,6), (7,0), (10,2) \Rightarrow \mu_2$

$\mu_1 = (2,3)$, $\mu_2 = (\frac{22}{3}, \frac{8}{3})$ $\Rightarrow$ Converge!

However, global minimum is

$(1,2), (3,4), (5,6) \Rightarrow \mu_1$

$(7,0), (10,2) \Rightarrow \mu_2$

$\mu_1 = (3,4)$, $\mu_2 = (8.5, 1)$

2. Programming Part

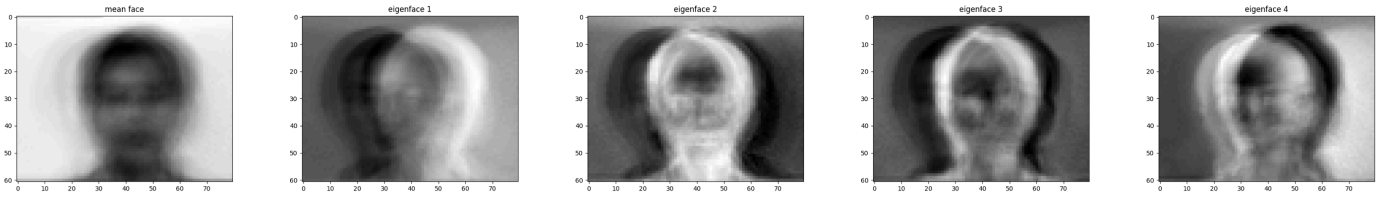
For the programming part, I consulted Copilot and ChatGPT.

2.1 Source code (Python)

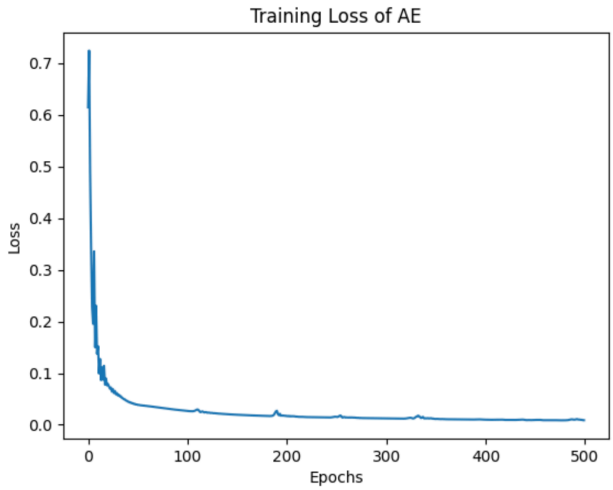
see hw4.py, src/pca.py, src/autoencoder.py

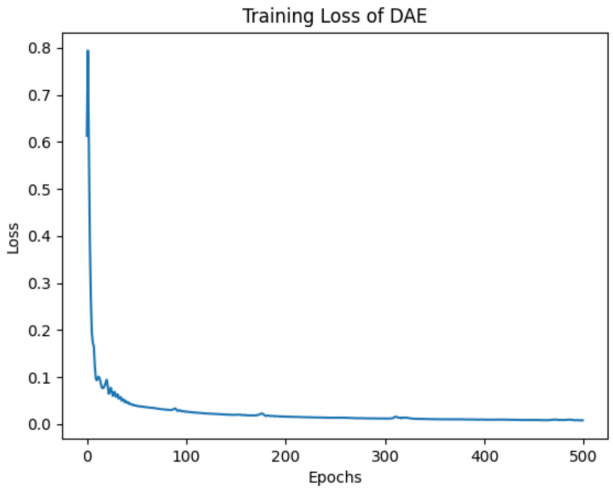
2.2 Report

(a)



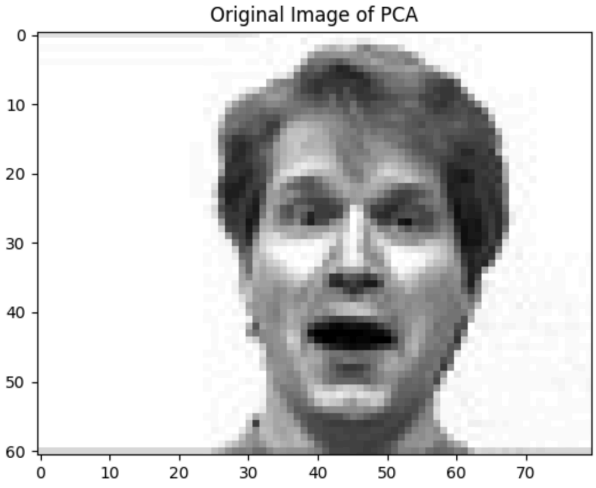
(b)

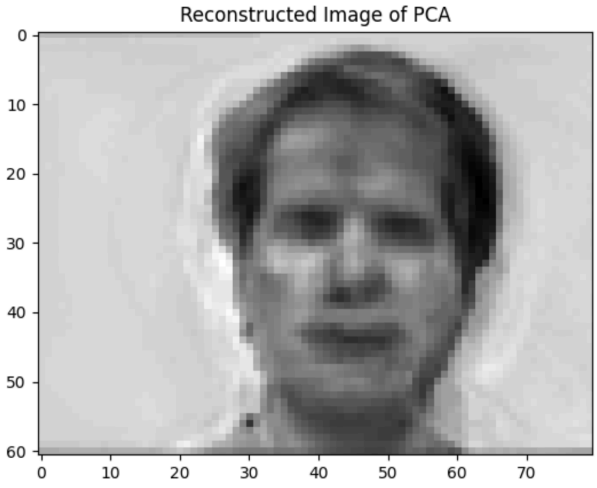




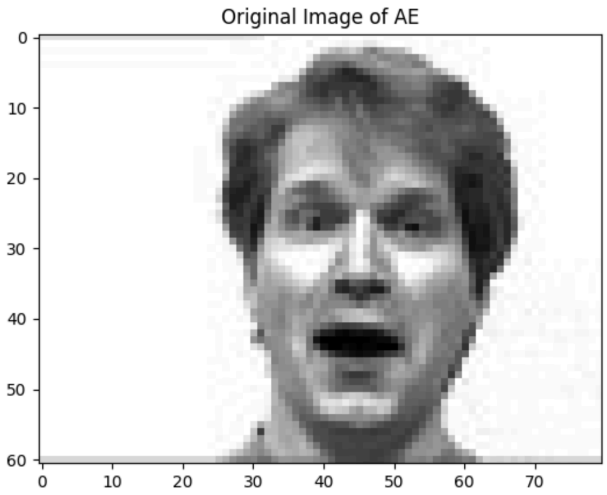
(c)

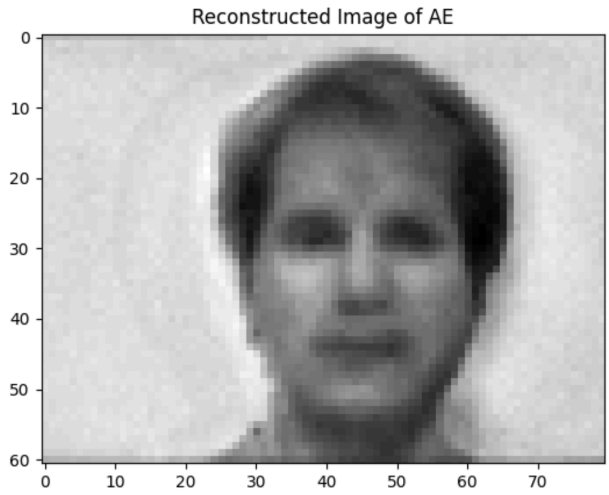
MSE with PCA: 0.010710469688056314



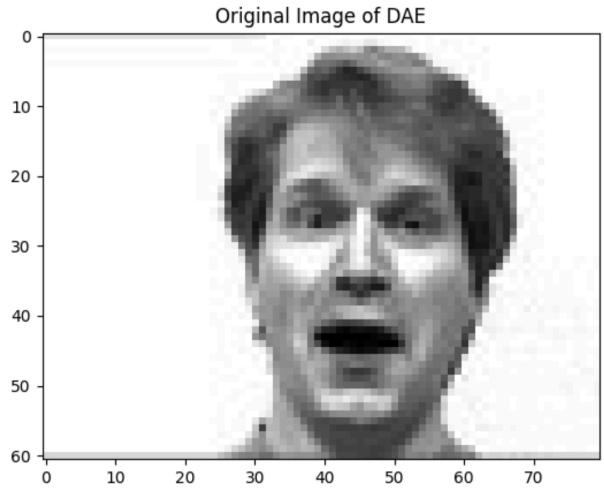


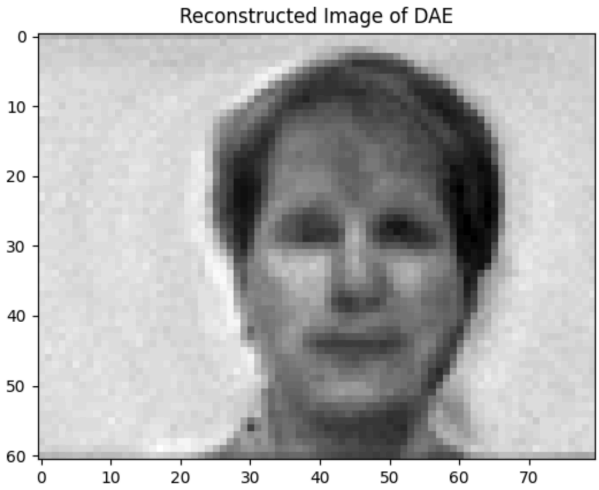
MSE with Autoencoder: 0.012450354173779488





MSE with DenoisingAutoencoder: 0.012580777783872379



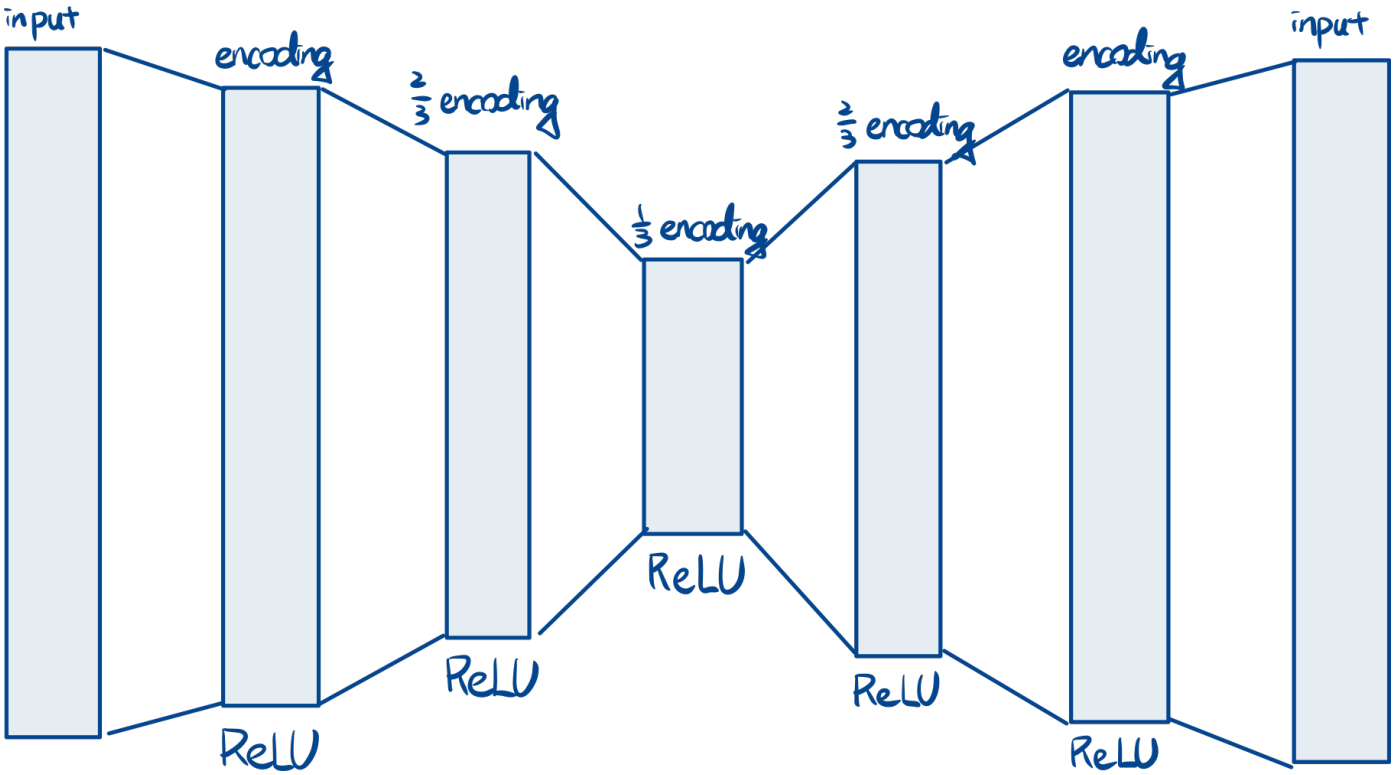


(d)

I have tried to make the denoising autoencoder deeper and shallower, but both result in lower accuracy. In my experiments, deeper model leads to 15% increase of reconstruction error, while shallower model leads to over 60% degradation.

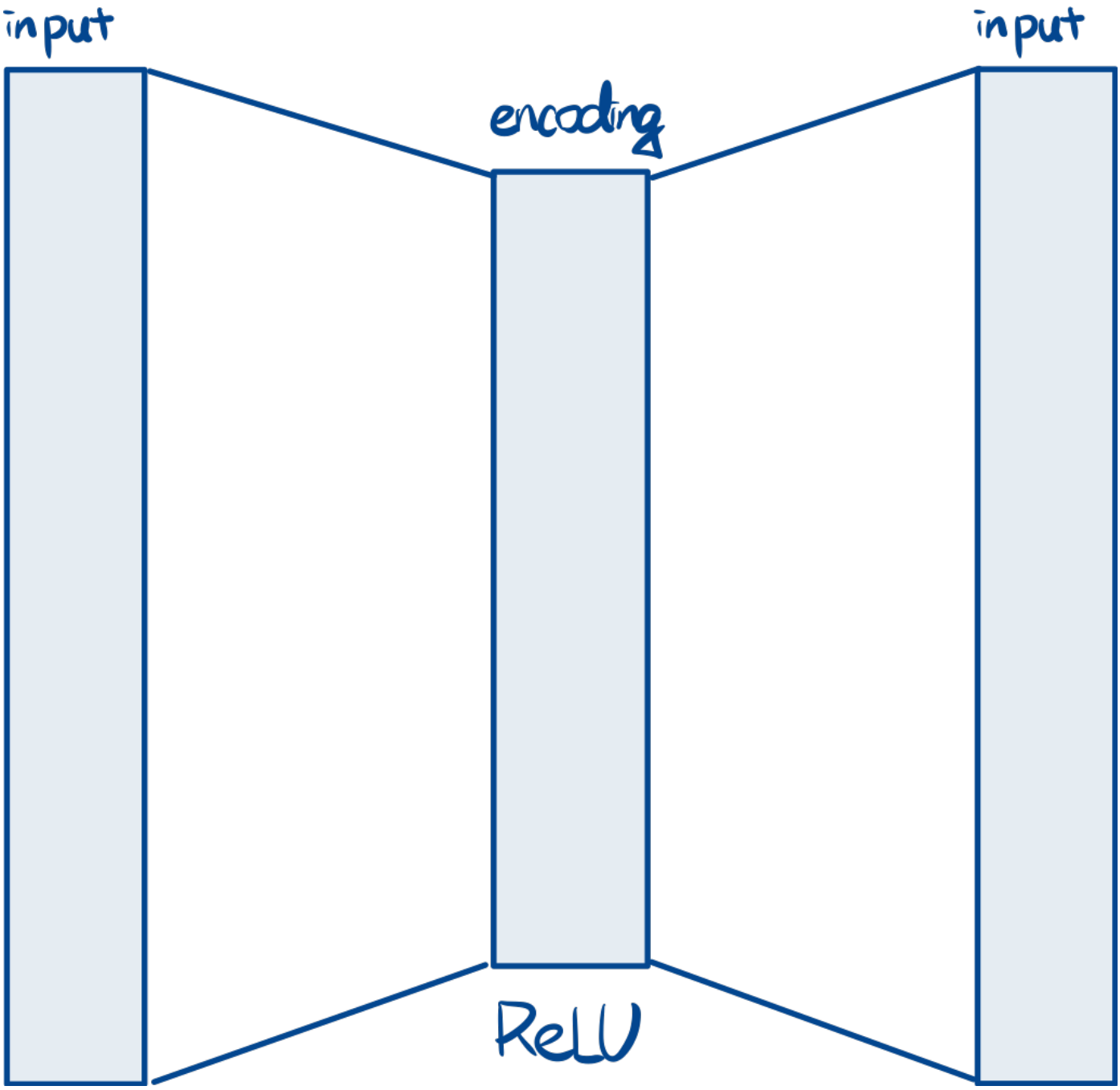
Deeper

Reconstruction Loss with DenoisingAutoencoder: 0.014202031412292088



Shallower

Reconstruction Loss with DenoisingAutoencoder: 0.02092486292693483

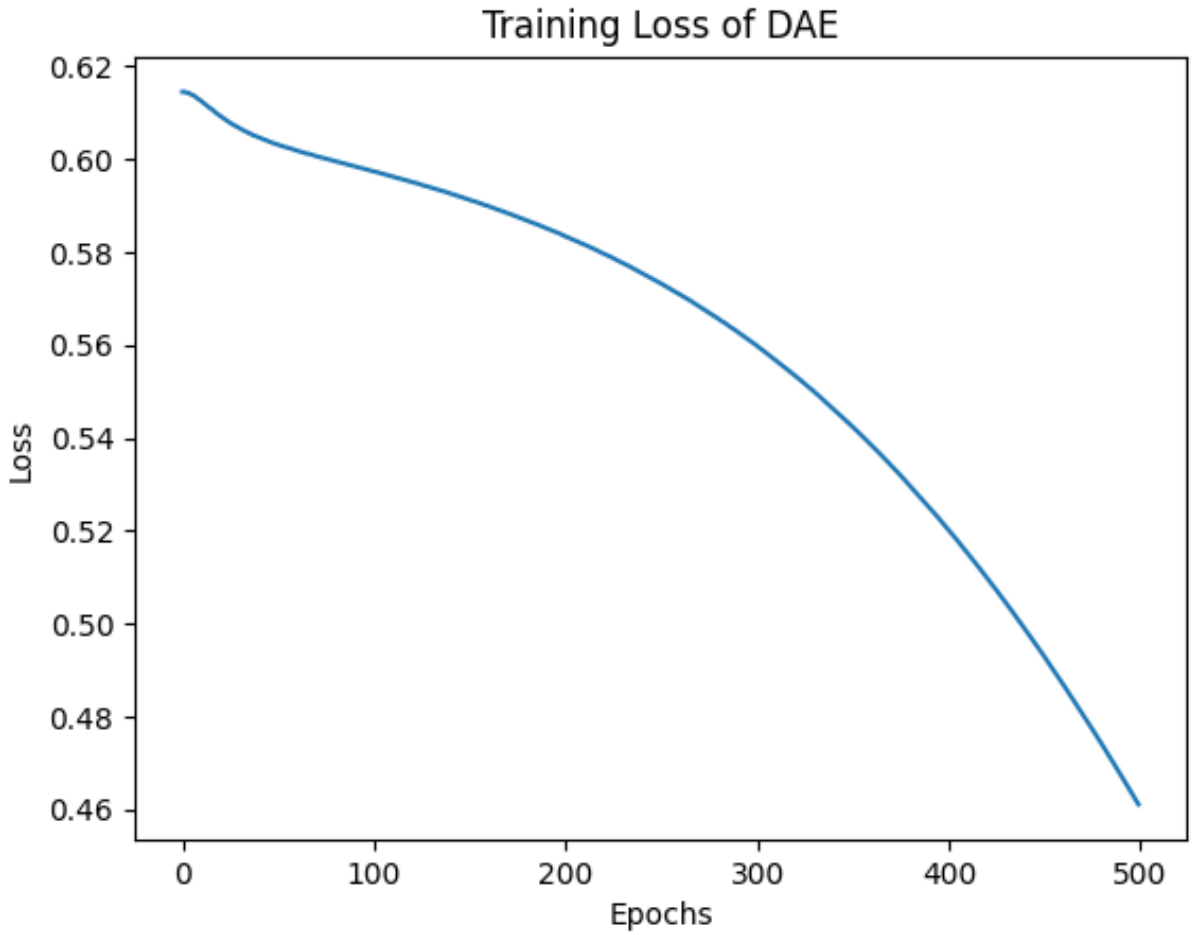


(e)

SGD: converge slower and doesn't converge after 500 epochs.

Acc from DenoisingAutoencoder: 0.8666666666666667

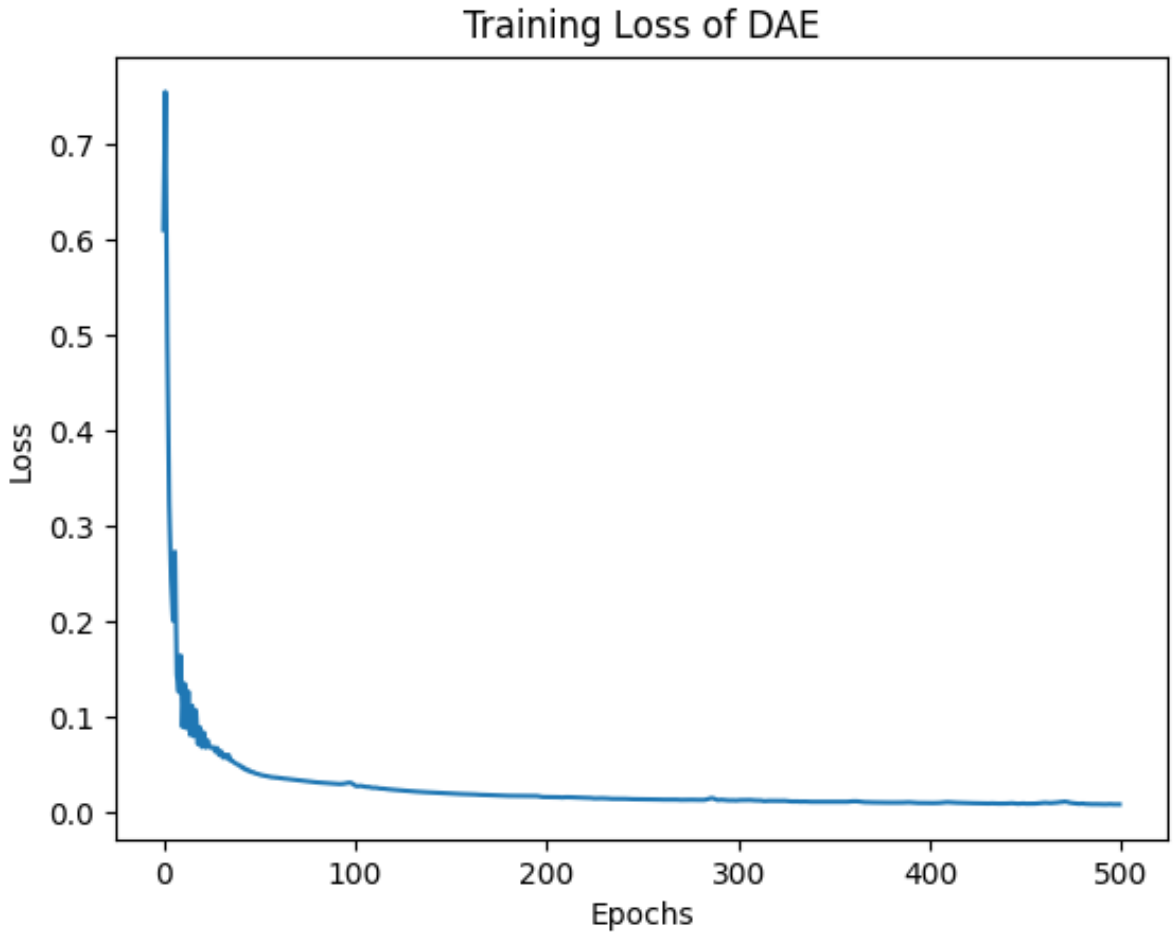
Reconstruction Loss with DenoisingAutoencoder: 0.5184948389173334



AdamW: curve similar to Adam, converges fast.

Acc from DenoisingAutoencoder: 0.9333333333333333

Reconstruction Loss with DenoisingAutoencoder: 0.013253448251617521



The performance of SGD is seriously degraded, while Adam and AdamW have similar performance (in the experiment, Adam is slightly better than AdamW).