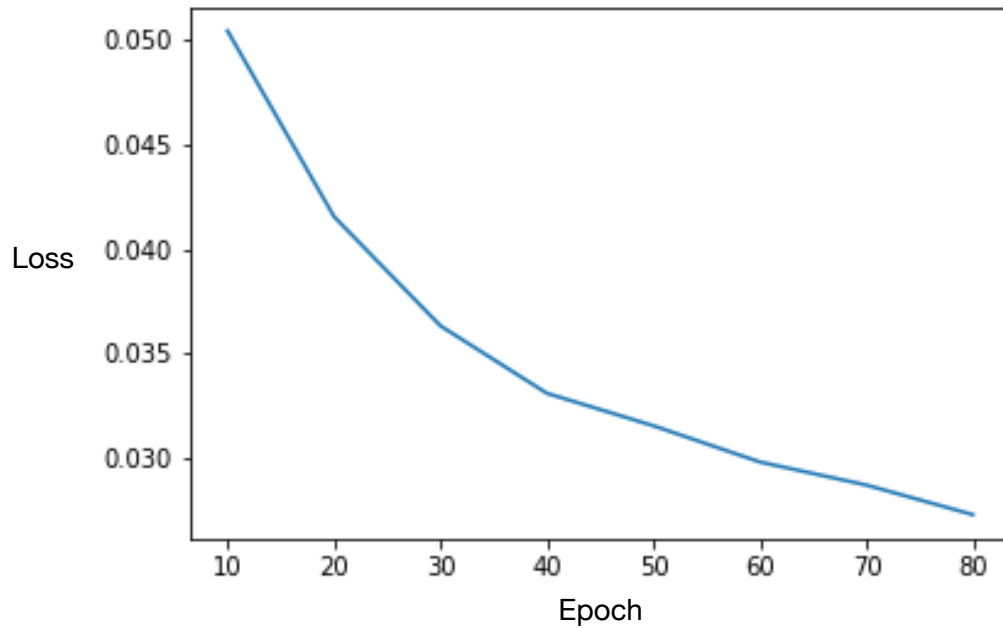


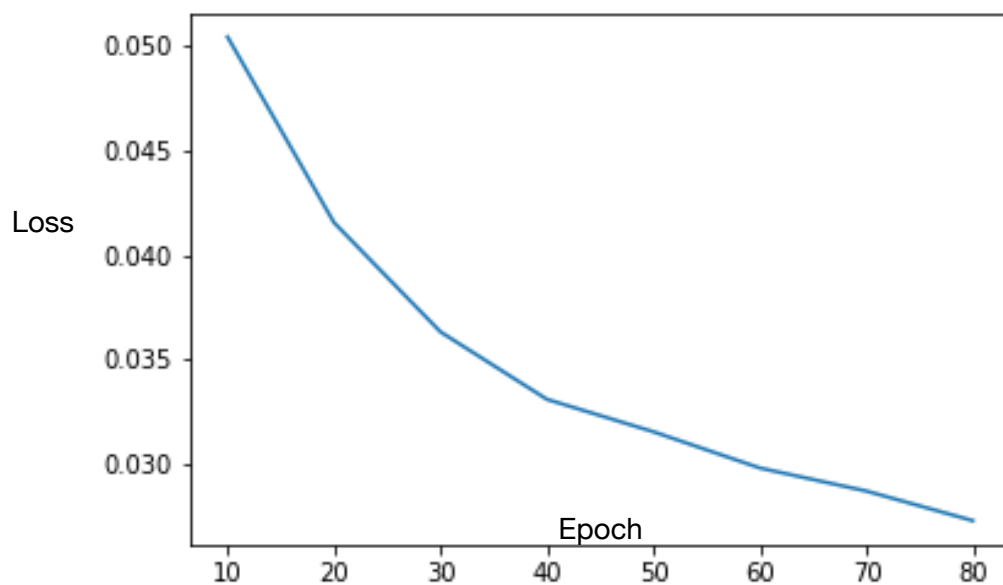
ML Hw4 Report

學號：B05705001 系級：資管四 姓名：黃意芹

1. 請使用不同的Autoencoder model，以及不同的降維方式(降到不同維度)，討論其reconstruction loss & public / private accuracy。（因此模型需要兩種，降維方法也需要兩種，但clustrering不用兩種。）



AUTOENCODER 1

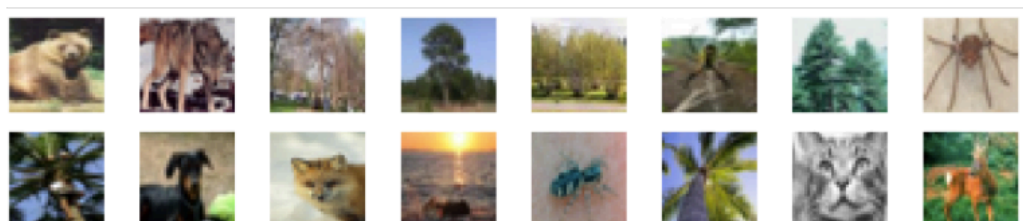


AUTOENCODER 2

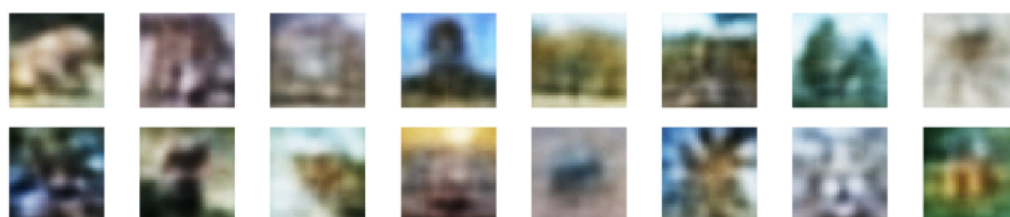
Public / Private Accuracy	Autoencoder 1	Autoencoder 2
PCA	0.53889 / 0.53207	0.72740 / 0.73412
TSNE	0.73148 / 0.71920	0.80740 / 0.80968

Autoencoder1的不同處為，model中的channel數量Autoencoder1開的比較大，Autoencoder2較小。而出來的latents維度Autoencoder1為200，Autoencoder2為128。Epoch數則都控制在80。結果發現Autoencoder2的loss收斂較快，而Accuracy的表現也是Autoencoder2較好。可見參數開大一點或latents維度多不代表表現就會好。

2. 從dataset選出2張圖，並貼上原圖以及經過autoencoder後reconstruct的圖片。

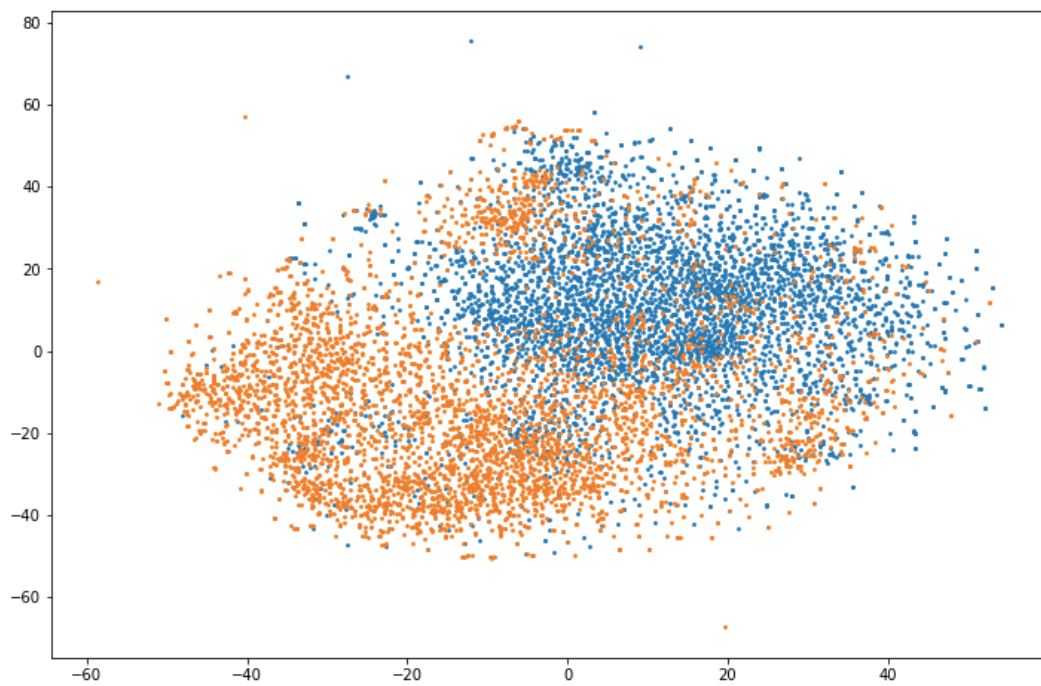


原圖

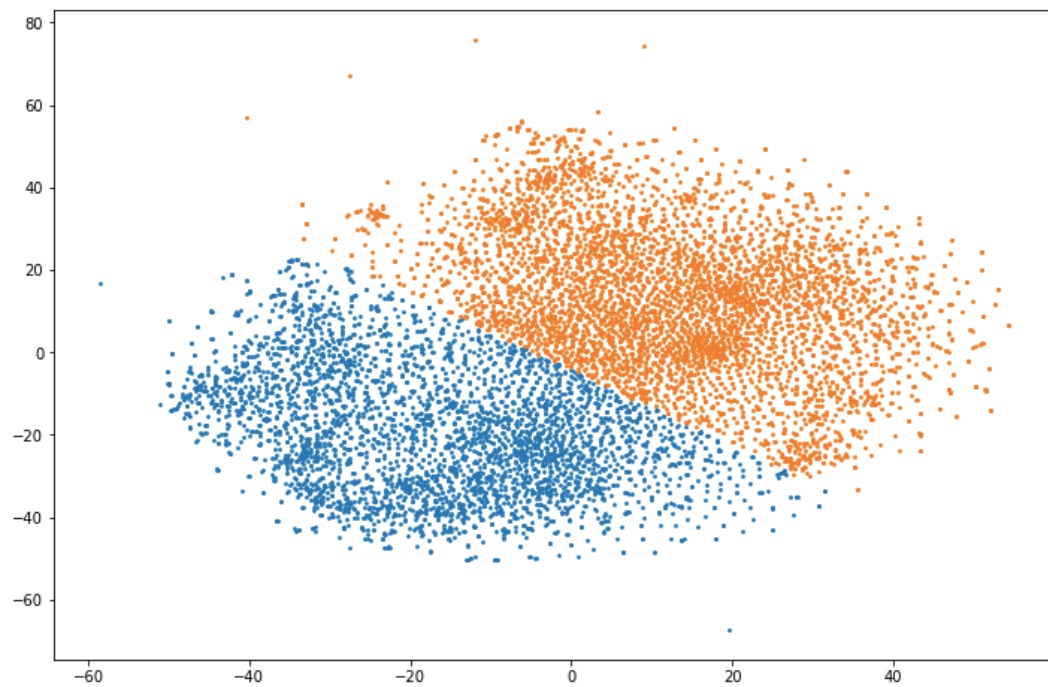


reconstruct後的圖

3. 在之後我們會給你dataset的label。請在二維平面上視覺化label的分佈。



TSNE降維後畫散佈圖，顏色對應真正的LABEL



TSNE降維後畫散佈圖，顏色對應到K-MEANS分群結果

4. Refer to math problem

ML HW4

$$1. S = \text{Cov}(X) = \frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})(x_i - \bar{x})^T \quad \bar{x} = [5.4 \ 8 \ 4.8]$$

$$= \begin{bmatrix} 12.04 & 0.5 & 3.28 \\ 0.5 & 12.2 & 2.9 \\ 3.28 & 2.9 & 8.16 \end{bmatrix}$$

Eigenvectors & Eigenvalue of S:

$$\lambda_3 = 5.49 \quad u_3 = \begin{bmatrix} 0.40 \\ 0.34 \\ -0.85 \end{bmatrix} \quad \lambda_2 = 11.63 \quad u_2 = \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix}$$

$$\lambda_1 = 15.30 \quad u_1 = \begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix}$$

(1) principal axes:

$$u_3 = \begin{bmatrix} 0.40 \\ 0.34 \\ -0.85 \end{bmatrix} \quad u_2 = \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix} \quad u_1 = \begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix}$$

$$(2) W = \begin{bmatrix} -0.62 & -0.59 & -0.52 \\ -0.68 & 0.73 & -0.03 \\ 0.40 & 0.34 & -0.85 \end{bmatrix}$$

$Z_i = W \times X_i$
(i-th principal component)

$$z_1 = \begin{bmatrix} -3.36 \\ 0.71 \\ 1.48 \end{bmatrix} \quad z_2 = \begin{bmatrix} -9.79 \\ 3.03 \\ -0.04 \end{bmatrix} \quad z_3 = \begin{bmatrix} -13.62 \\ 6.53 \\ 2.42 \end{bmatrix} \quad z_4 = \begin{bmatrix} -7.94 \\ 5.06 \\ 1.16 \end{bmatrix}$$

$$z_5 = \begin{bmatrix} -12.37 \\ 6.84 \\ -5.02 \end{bmatrix} \quad z_6 = \begin{bmatrix} -7.19 \\ -1.84 \\ -3.30 \end{bmatrix} \quad z_7 = \begin{bmatrix} -14.96 \\ -0.47 \\ 1.37 \end{bmatrix} \quad z_8 = \begin{bmatrix} -7.08 \\ 3.81 \\ -3.05 \end{bmatrix}$$

$$z_9 = \begin{bmatrix} -12.86 \\ -3.95 \\ -0.97 \end{bmatrix} \quad z_{10} = \begin{bmatrix} -16.30 \\ 1.11 \\ -1.75 \end{bmatrix}$$

$$(3) \text{ average reconstruction error} = \frac{1}{10} \sum_{i=1}^{10} (x_i - y_i)^2 = 6.064$$

$$y_i = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \cdot \begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

2. Let i th row in A be a_i

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \quad A^T = \begin{bmatrix} | & | & \dots & | \\ a_1 & a_2 & \dots & a_m \\ | & | & \dots & | \end{bmatrix}$$

Let $B = AA^T = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & \dots & \dots & b_{nn} \end{bmatrix}$ $\because b_{ij} = a_i \cdot a_j = a_j \cdot a_i = b_{ji}$ By 内积对称
 $\therefore B$ is symmetric

Let i th row in A^T be a'_i

$$A^T = \begin{bmatrix} a'_1 \\ \vdots \\ a'_n \end{bmatrix} \quad A = \begin{bmatrix} | & \dots & | \\ a'_1 & \dots & a'_n \\ | & \dots & | \end{bmatrix}$$

$\in \mathbb{R}^{n \times m}$

Let $B' = A^T A = \begin{bmatrix} b'_{11} & \dots & b'_{1m} \\ \vdots & \ddots & \vdots \\ b'_{m1} & \dots & b'_{mm} \end{bmatrix}$ $\because b'_{ij} = a'_i \cdot a'_j = a'_j \cdot a'_i = b'_{ji}$
 $\therefore B'$ is symmetric

Therefore, AA^T and $A^T A$ are symmetric

positive semi-definite: $x^T \Sigma x \geq 0 \quad \forall x \in \mathbb{R}^n$, where $\Sigma \in \mathbb{R}^{n \times n}$

Let $B = AA^T$ Let $C = A^T x$

$$x^T A A^T x = A^T x A^T x = C C$$

Let $D = C C$, $d_{ij} = C_i \cdot C_i \geq 0$ By 内积正定性

Let $E = A^T A$ Let $F = A x$

$$x^T A^T A x = A x A x = F F$$

Let $G = F F$, $g_{ij} = f_i \cdot f_i \geq 0$

Therefore, AA^T and $A^T A$ are positive semi-definite

$$AA^T \sim A^T A$$

$$\because QAA^T = QA^T A = AQA^T = AA^T Q$$

$$\xrightarrow{\times Q^{-1}} QAA^T Q^{-1} = AA^T Q Q^{-1} = AA^T \quad \therefore AA^T \sim A^T A \quad \text{--- ①}$$

$$\text{Let } B = AA^T, \quad C = A^T A$$

$$\begin{aligned} \det(B) &= \det(Q \times C \times Q^{-1}) = \det(Q) \cdot \det(C) \cdot \det(Q^{-1}) \\ &= \det(Q) \cdot \det(Q^{-1}) \cdot \det(C) = \det(I) \cdot \det(C) = \det(C) \quad \text{--- ②} \end{aligned}$$

$$B - tI_n \sim Q(B - tI_n)Q^{-1} = QBQ^{-1} - tQI_nQ^{-1} = C - tI_n$$

$$\text{By ①②, } \because B - tI_n \sim C - tI_n \quad \therefore \det(B - tI_n) = \det(C - tI_n)$$

Therefore when $\det(B - \lambda I) = 0$, $\det(C - \lambda I) = 0$.

$\rightarrow B$ and C have the same eigenvalues.

$$AA^T x = \lambda x \xrightarrow{\times A^T} (A^T A)A^T x = \lambda A^T x \rightarrow \lambda = A^T A = \|A\|^2$$

if $A \neq 0_{R^{m \times n}}$, $\|A\|^2 > 0$ $\therefore B$ and C have non-zero eigenvalue

(b)

$$\text{Let } x = (x_1, x_2, \dots, x_n) \in R^n$$

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T$$

Let $M = (x - \mu)(x - \mu)^T$, M is a symmetric positive semi-definite matrix (by 2(a)) $\therefore z^T M z \geq 0 \quad \forall z \in R^m$

$$z^T M z \geq 0 \rightarrow E(z^T M z) \geq 0 \rightarrow z^T E(M) z \geq 0$$

$$\text{其中 } E(M) = E[(x - \mu)(x - \mu)^T] = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T = \Sigma$$

$$\therefore z^T \Sigma z \geq 0$$

Therefore, Σ is a symmetric positive semi-definite matrix.

(C) Take $X = [x_1 \ x_2 \ \dots \ x_n]$ s.t. $\Sigma = \frac{1}{N} X X^T = U \Lambda U^T$

$$\text{Trace}(\Phi^T \Sigma \Phi) = \frac{1}{N} \text{Tr}(\Phi^T X X^T \Phi) = \frac{1}{N} \|\Phi^T X\|_F^2$$

$$= \frac{1}{N} \sum_{i=1}^N \|\Phi^T x_i\|^2 = \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i^{(S)}\|^2$$

$$\rightarrow 0 \leq \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i^{(S)}\|^2 \leq \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i^{(PCA)}\|^2$$

3.

Gradient Decent:

$$g_t \leftarrow g_{t-1} - \eta \frac{\partial L(g)}{\partial g}$$

We hope $\alpha_t f_t(x)$ and $-\eta \frac{\partial L(g)}{\partial g}$ have the same direction,

Boosting

so the loss can be reduced,

$$g_t(x) \leftarrow g_{t-1}(x) + \alpha_t f_t(x)$$

$$L = \sum_{i=1}^n \exp\left(\frac{1}{K-1} \sum_{k \neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right)$$

$$\frac{\partial L}{\partial g} = \sum_{i=1}^n \exp\left(\frac{1}{K-1} \sum_{k \neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right) \times \frac{1}{K-1} \sum_{k \neq \hat{y}_i} (1 - g_T^{\hat{y}_i}(x_i))$$

We want to find $f_t(x)$ maximizing \sum , so $f_t(x)$ and $\eta \frac{\partial L}{\partial g}$ are likely to

$$\sum_{i=1}^n \exp\left(\frac{1}{K-1} \sum_{k \neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right) \times \frac{1}{K-1} \sum_{k \neq \hat{y}_i} (1 - g_T^{\hat{y}_i}(x_i)) \cdot f_t(x_i)$$

Then we want to find α_t minimizing $L(g)$

be in the same direction

$$L(g) = \sum_{i=1}^n \exp\left(\frac{1}{K-1} \sum_{k \neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right)$$

$$\frac{\partial L}{\partial \alpha_t} = \sum_{i=1}^n \exp\left(\frac{1}{K-1} \sum_{k \neq \hat{y}_i} \left(\sum_{t=1}^T \alpha_t^k f_t(x) - \alpha_t^{\hat{y}_i} f_t(x)\right)\right) \times \frac{1}{K-1} \sum_{k \neq \hat{y}_i} \sum_{t=1}^T f_t(x) - \alpha_t^{\hat{y}_i} f_t(x) = 0$$

In this way, we can minimize the loss function

update α_t^k and f_t to