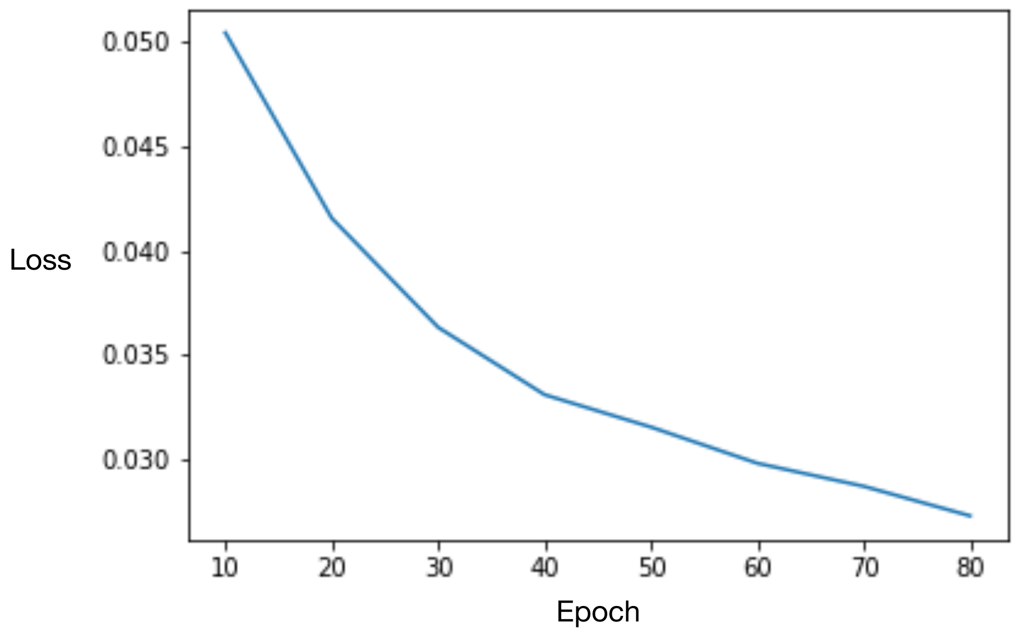
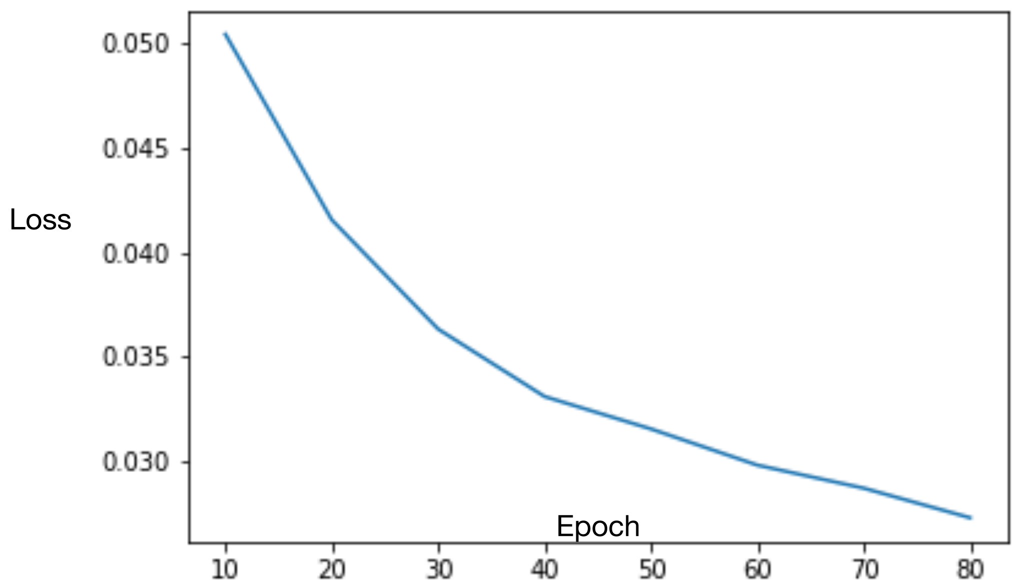
ML Hw4 Report

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



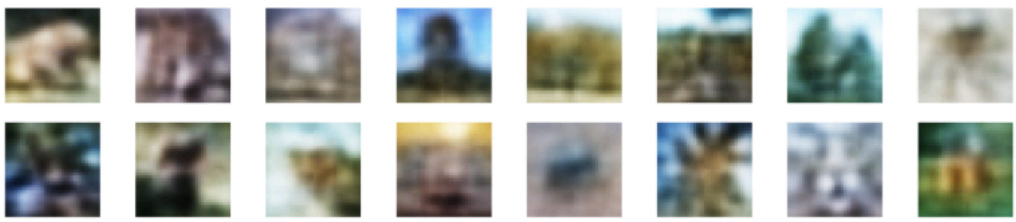
AUTOENCODER 2

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。

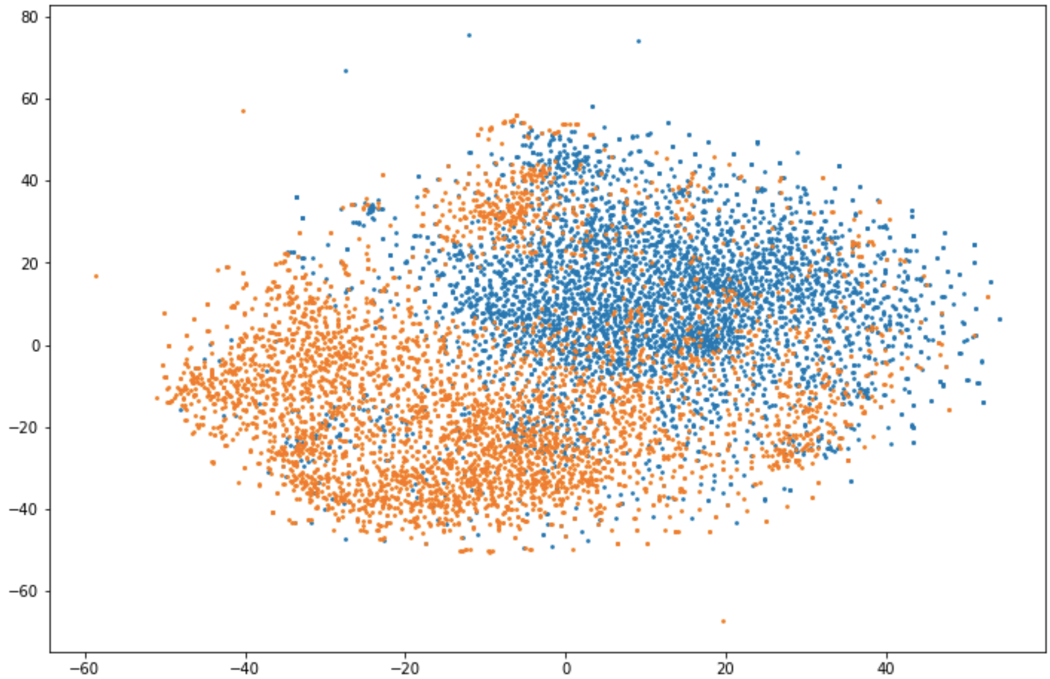
2. 從dataset選出2張圖，並貼上原圖以及經過autoencoder後reconstruct的圖片。

原圖

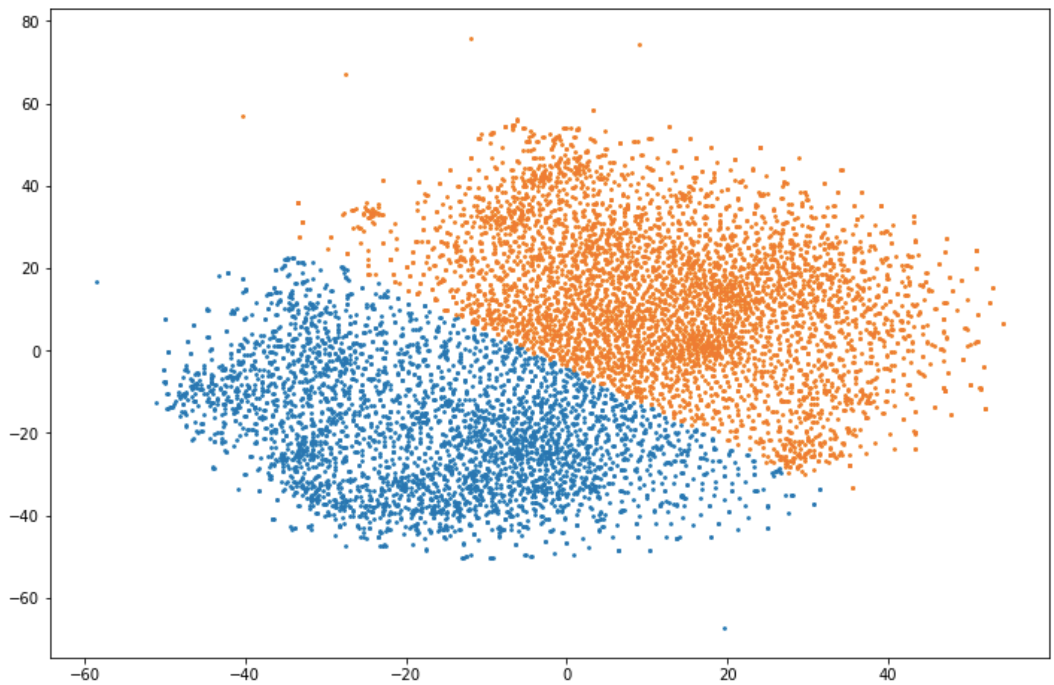


reconstruct後的圖

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

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

Eigenvectors & Eigenvalue of S:

$$\lambda_3 = 5.47 \quad u_3 = \begin{bmatrix} 0.40 \\ 0.34 \\ -0.85 \end{bmatrix} \quad \lambda_2 = 11.63 \quad u_2 = \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix}$$

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

(1) pricipal 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) $W = \begin{bmatrix} -0.62 & -0.59 & -0.52 \\ -0.68 & 0.73 & -0.03 \\ 0.40 & 0.34 & -0.85 \end{bmatrix}$ $\qquad z_i = W \times x_i$

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

$$y_i = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \cdot \begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

2. Let ith row in A be a_i

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

By 内积对称

Let $B = AA^T = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ \vdots & & & \vdots \\ b_{n1} & - & - & b_{nn} \end{bmatrix}$ $\because b_{ij} = a_i \cdot a_j = a_j \cdot a_i = b_{ji}$

$\therefore$ B is symmetric

Let ith row in A^T be a_i'

$$A^T = \begin{bmatrix} - a_1' - \\ \vdots \\ - a_n' - \end{bmatrix} \in R^{n \times m} \quad A = \begin{bmatrix} a_1' & \cdots & a_n' \end{bmatrix}$$

Let $B' = A^T A = \begin{bmatrix} b_{11}' & \cdots & b_{1m}' \\ \vdots & \ddots & \vdots \\ b_{m1}' & \cdots & b_{mm}' \end{bmatrix}$ $\because b_{ij}' = a_i' \cdot a_j' = a_j' \cdot a_i' = b_{ji}'$

$\therefore$ B' is symmetric

Therefore, AA^T and A^TA are symmetric

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

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

$x^T A A^T x = A^T x \, A^T x = CC$

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

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

$x^T A^T A x = Ax \, Ax = FF$

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

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

$AA^T \sim A^TA$

$\because QAA^T = QA^TA = AQ^TA = AA^TQ$

$\xrightarrow{\times Q^{-1}} QAA^TQ^{-1} = AA^TQQ^{-1} = AA^T \quad \therefore AA^T \sim A^TA$ —— ①

Let $B = AA^T$, $C = A^TA$

$\det(B) = \det(Q \times C \times Q^{-1}) = \det(Q) \cdot \det(C) \cdot \det(Q^{-1})$

$= \det(Q) \cdot \det(Q^{-1}) \cdot \det(C) = \det(I) \cdot \det(C) = \det(C)$ — ②

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

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.

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

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

(b)

Let $x = (x_1, x_2, \cdots x_n) \in R^m$

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

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

$z^TMz \geq 0 \rightarrow E(z^TMz) \geq 0 \rightarrow z^TE(M)z \geq 0$

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 $X = [x_1 \ x_2 \cdots x_n]$ s.t. $\Sigma = \frac{1}{N} X X^T = U \Lambda U^T$

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

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

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

3、 For kth class, first we want to find f_t s.t.

$$L(g_t^1 \cdots g_t^k) < L(g_{t+1}^1 \cdots g_{t+1}^k)$$

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

$$= \sum_{i=1}^{n} \exp\left(\frac{1}{k-1}\left(\sum_{k \neq \hat{y}_i} \sum_{t=1}^{T} a_t^k f_t(x) - \sum_{t=1}^{T} a_t^{\hat{y}_i} f_t(x)\right)\right)$$

$$= \sum_{\substack{i=1 \\ (k \neq y_i)}}^{n} \exp\left(\frac{1}{k-1} \sum_{k \neq \hat{y}_i} g_{t-1}^k(x_i) + \frac{a_t^k f_t(x)}{k-1} - g_t^{\hat{y}_i}(x_i)\right) +$$

$$\sum_{\substack{i=1 \\ (k = y_i)}}^{n} \exp\left(\frac{1}{k-1} \sum_{k \neq \hat{y}_i} g_{t-1}^k(x_i) - a_t^k f_t(x) - g_t^{\hat{y}_i}(x_i)\right)$$

$$\Delta L = L(g_T^k) - L(g_{t-1})^k \qquad \because a_t^k \ll 1 \quad \approx \frac{a_t^k f_t(x)}{k-1}$$

$$= \sum_{\substack{i=1 \\ (k = y_i)}}^{n} \exp\left(\frac{1}{k-1} \sum_{k \neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right)\left(\exp\left(\frac{a_t^k f_t(x_i)}{k-1}\right) - 1\right) +$$

$$\sum_{\substack{i=1 \\ (k = y_i)}}^{n} \exp\left(\frac{1}{k-1} \sum_{k \neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x)\right)\left(\exp\left(-a_t^k f_t(x_i) - 1\right)\right)$$

$$\because a_t^k \ll 1 \quad \approx -a_t^k f_t(x_i)$$

$$\Delta L = \sum_{i=1}^{n} \exp\left(\frac{1}{k-1} \sum_{k \neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right) \cdot \frac{a_t^k f_t(k)}{k-1}$$

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

$$\frac{\partial L}{\partial a_t} = \frac{f_t(k)}{k-1} \sum_{(k \neq y_i)} \exp\left(\frac{1}{k-1} \sum_{k \neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right)$$

$$- f_t(x_i) \sum_{(k = y_i)} \exp\left(\frac{1}{k-1} \sum_{k \neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right)$$

Therefore, it's like gradient decent

we can update $f_t \leftarrow f_{t-1} - \eta \frac{\partial L}{\partial a_t}$

To find a_t minimizing L

$$\frac{\partial L}{\partial a_t} = \sum_{k \neq y_i} u_t^n e^{a_t} \cdot \frac{1}{k-1} - \sum_{k = y_i} u_t^n e^{-a_t} = 0$$

$$\rightarrow Z_t \varepsilon_t e^{a_t} \frac{1}{k-1} - Z_t (1-\varepsilon_t) e^{-a_t} = 0$$

$$\rightarrow e^{2a_t} = \left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) k-1$$

$$\rightarrow a_t = \ln\left(\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t} k-1}\right)$$ ※