學號：R08725021 系級： 資管碩一 姓名：王鼎元

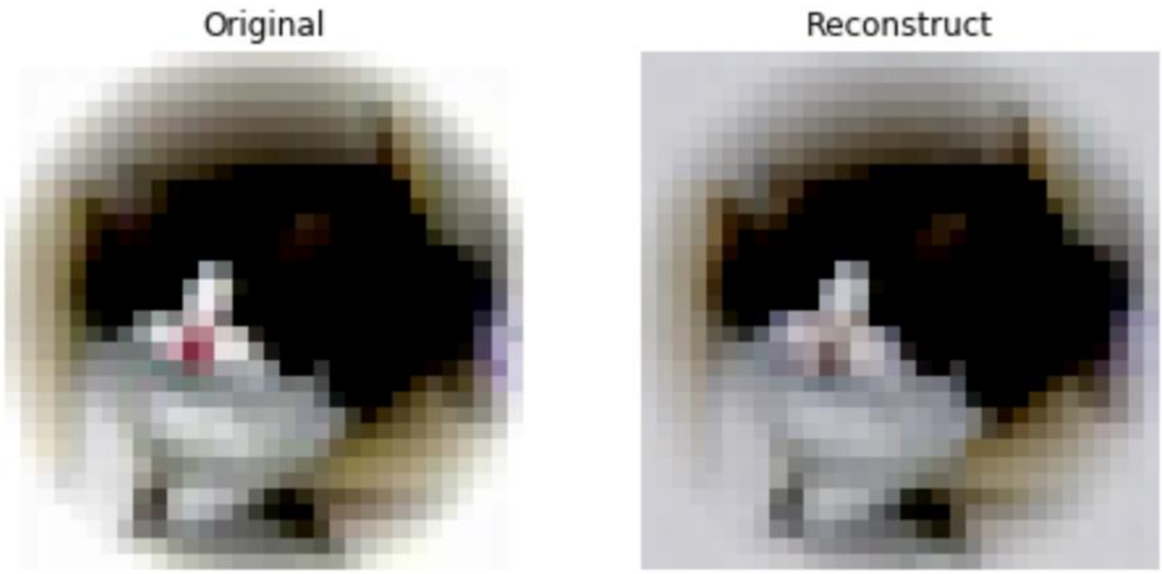
1. (1%) 請使用不同的Autoencoder model，以及不同的降維方式(降到不同維度)，討論其reconstruction loss & public / private accuracy。（因此模型需要兩種，降維方法也需要兩種，但clustrering不用兩種。）

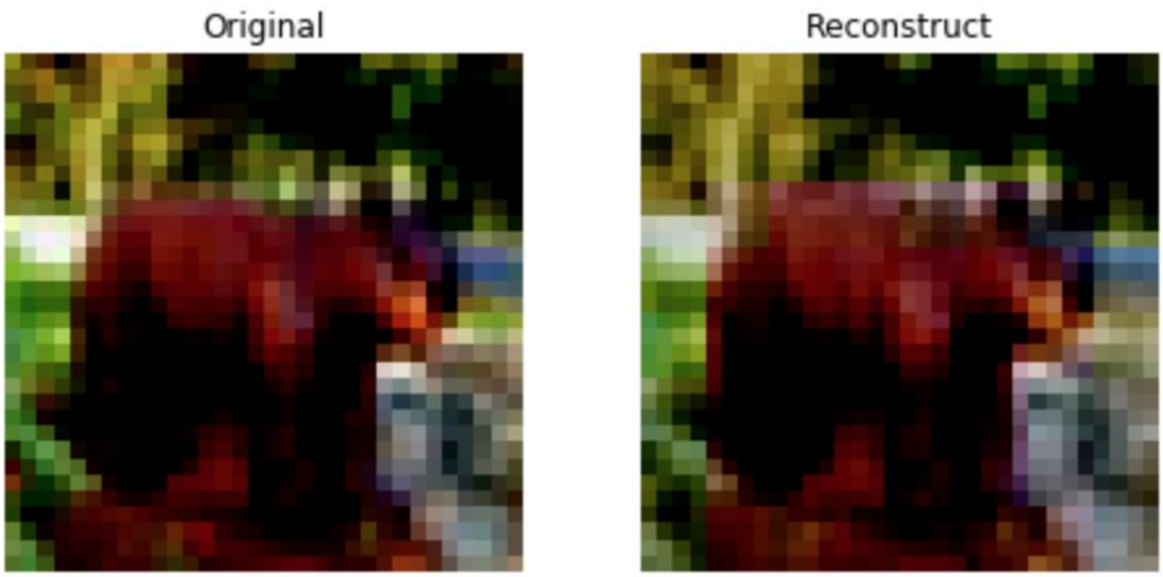
Model 1 是先用兩層的 autoencoder 取出 (16, 8, 8) 的 latent，transform 成 16x8x8=1024 維的向量，再用 whiten PCA 降維到 32 維，最後用 Kmeans 分群；Model 2 是先用三層的 autoencoder 取出 (32, 4, 4) 的 latent，transform 成 32x4x4=512 維的向量，用 whiten PCA 降維到 32 維，再用 tSNE 降維到 8 維，最後用 Kmeans 分群。成果如下：

	reconstruction loss	public accuracy	private accuracy
Model 1	139.7827	0.79666	0.78476
Model 2	194.2429	0.80888	0.81666

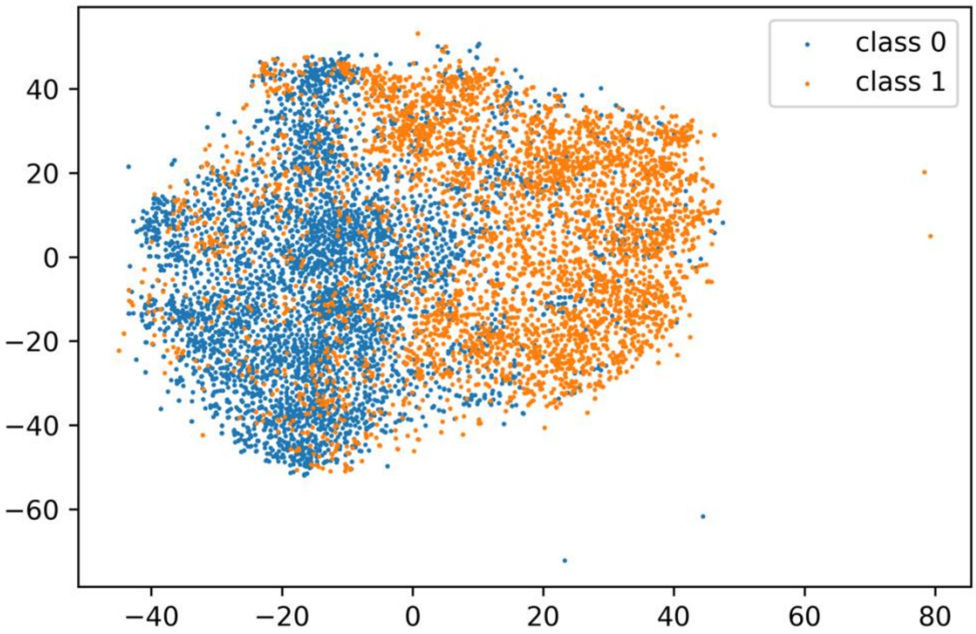
由此可知，autoencoder 的層數並不是越深越好，可能在壓縮過程中損失太多特徵；PCA+tSNE 壓縮完後會讓不同群的資料比較分開，做分群比較容易有好的結果。

2. (1%) 從dataset選出2張圖，並貼上原圖以及經過autoencoder後reconstruct的圖片。





3. (1%) 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。



4. (3%)Refer to math problem
https://drive.google.com/file/d/1e_IDAV2yv0YEhIuVWpDdaH4Pzz5s1p2P/view?fbclid=IwAR0tO9NRxK9JZeUDNdawNuSbGTvqI7niuMX3Kkk9arauC8O6p6iJc7oMz84

Q1

(a) ¶

- 先求出 μ, Σ

$$\mu = \frac{1}{10}([1\ 2\ 3]^T + [4\ 8\ 5]^T + [3\ 12\ 9]^T + \ldots + [10\ 11\ 7]^T) = [5.4\ 8\ 4.8]^T$$

$$\Sigma = \frac{1}{10}\Sigma_{i=1}^{10}(x_i - \mu)(x_i - \mu)^T = \begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix}$$

- 令 $det(\Sigma - \lambda I) = 0$, 以求 eigenvalue λ 與 eigenvector v

$$\lambda = 6.08,\ 12.92,\ 17.00$$

$$v = \begin{bmatrix} 0.40 & -0.68 & -0.62 \\ 0.34 & 0.73 & -0.59 \\ -0.85 & -0.03 & -0.52 \end{bmatrix}$$

- 依 λ 的大小將 v 做降序排列, 最大者即是 principal axis

$$v_{principal} = \begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix}$$

(b)

- 經降維後的資料 z_i 可以由 $z_i = v^T \cdot x_i$ 算出

$$z = v^T \cdot x = \begin{bmatrix} -0.62 & -0.68 & 0.40 \\ -0.59 & 0.73 & 0.34 \\ -0.52 & -0.03 & -0.85 \end{bmatrix}^T \cdot \begin{bmatrix} 1 & 4 & 3 & 1 & 5 & 7 & 9 & 3 & 11 & 10 \\ 2 & 8 & 12 & 8 & 14 & 4 & 8 & 8 & 5 & 11 \\ 3 & 5 & 9 & 5 & 2 & 1 & 9 & 1 & 6 & 7 \end{bmatrix}$$

得 $z = \begin{bmatrix} -3.36 & -9.79 & -13.62 & -7.94 & -12.37 & -7.19 & -14.96 & -7.08 & -12.86 & -16.30 \end{bmatrix}$

(c)

- Reconstruction error 需要用 z_i 回求原本高維度的 $\bar{x}_i$, 再與 x_i 做 mean squared error
- 求壓縮到二維的 z

$$z = v^T \cdot x = \begin{bmatrix} -0.62 & -0.68 \\ -0.59 & 0.73 \\ -0.52 & -0.03 \end{bmatrix}^T \cdot \begin{bmatrix} 1 & 4 & 3 & 1 & 5 & 7 & 9 & 3 & 11 & 10 \\ 2 & 8 & 12 & 8 & 14 & 4 & 8 & 8 & 5 & 11 \\ 3 & 5 & 9 & 5 & 2 & 1 & 9 & 1 & 6 & 7 \end{bmatrix}$$

得 $z = \begin{bmatrix} -3.36 & -9.79 & -13.62 & -7.94 & -12.37 & -7.19 & -14.96 & -7.08 & -12.86 & -16.30 \\ 0.71 & 3.03 & 6.53 & 5.06 & 6.84 & -1.84 & -0.47 & 3.81 & -3.95 & 1.11 \end{bmatrix}$

- 求 $\bar{x}$

$$\bar{x} = v \cdot z = \begin{bmatrix} -0.62 & -0.68 \\ -0.59 & 0.73 \\ -0.52 & -0.03 \end{bmatrix} \cdot \begin{bmatrix} -3.36 & -9.79 & -13.62 & -7.94 & -12.37 & -7.19 & -14.96 & -7.08 & -12.86 & -16.30 \\ 0.71 & 3.03 & 6.53 & 5.06 & 6.84 & -1.84 & -0.47 & 3.81 & -3.95 & 1.11 \end{bmatrix}$$

得 $\bar{x} = \begin{bmatrix} 1.60 & 3.98 & 3.97 & 1.46 & 2.99 & 5.68 & 9.55 & 1.78 & 10.61 & 9.30 \\ 2.50 & 7.99 & 12.82 & 8.39 & 12.30 & 2.89 & 8.46 & 6.97 & 4.67 & 10.41 \\ 1.74 & 5.03 & 6.94 & 4.01 & 6.28 & 3.81 & 7.82 & 3.60 & 6.83 & 8.49 \end{bmatrix}$

- 再用 $\bar{x}$ 跟 x 算 reconstruction error

$$MSE_{recons} = \frac{1}{10}\Sigma_{i=1}^{10}(x_i - \bar{x}_i)^2$$

得 $MSE_{recons} = 6.06$

Q2

(a)

- 證: AA^T 與 A^TA 都是 symmetric matrix

$$(AA^T)^T = (A^T)^TA^T = AA^T$$
$$(A^TA)^T = A^T(A^T)^T = A^TA$$
$$\text{兩者的轉置矩陣都是自己, 故均為對稱矩陣}$$

- 證: AA^T 與 A^TA 都是 semi-definite matrix

$$X^T(AA^T)X = (A^TX)^TA^TX > 0$$
$$X^T(A^TA)X = (AX)^TAX > 0$$
$$\text{兩式成立, 故均為半正定矩陣}$$

- 證: AA^T 與 A^TA 有相同的非零 eigenvalue

$$(AA^T)v = \lambda v$$
$$A^T(AA^T)v = A^T\lambda v$$
$$(A^TA)(A^T)v = \lambda(A^Tv)$$
$$\text{令 } v' = (A^T)v$$
$$\text{得 } (A^TA)v' = \lambda v'$$
$$\text{由上可知, } AA^T \text{ 與 } A^TA \text{ 有相同的非零 } eigenvalue\, \lambda$$

(b)

- Covariance matrix 的公式為 $\Sigma = \frac{1}{n}\Sigma_i(x_i - \mu)(x_i - \mu)^T$, 故只要讓 $\mu = 0$ 即可

$$\text{令 } x = [x1\ x2\ x3\ x4] = \begin{bmatrix} 0 & 2 & -4 & 1 \\ 1 & -7 & 10 & -4 \\ 2 & -3 & -5 & 6 \end{bmatrix}$$

(c)

- 已知 $\Sigma = \frac{1}{N}XX^T = U\Lambda U^T$ 其中 U 為 orthogonal matrix, Λ 為 diagonal matrix
- 又知如果 A 為 diagonal matrix

$$Tr(A) = ||A||$$

- 故

$$\begin{aligned} & Tr(\phi^T\Sigma\phi) \\ = & \frac{1}{N}Tr(\phi^TXX^T\phi) \\ = & \frac{1}{N}||Q^TX||^2 \\ = & \frac{1}{N}\Sigma_i||Q^Tx_i||^2 \end{aligned}$$

- 令 $\hat{x}_i^{(s)} = Q^Tx_i$

$$\begin{aligned} & \frac{1}{N}\Sigma_i||Q^Tx_i||^2 \\ = & \frac{1}{N}\Sigma_i||\hat{x}_i^{(s)}||^2 \leq \frac{1}{N}\Sigma_i||\hat{x}_i^{(PCA)}||^2 \end{aligned}$$

- 又因 Σ 是 semi-definite matrix

$$0 \leq \frac{1}{N}\Sigma_i||\hat{x}_i^{(s)}||^2 \leq \frac{1}{N}\Sigma_i||\hat{x}_i^{(PCA)}||^2$$