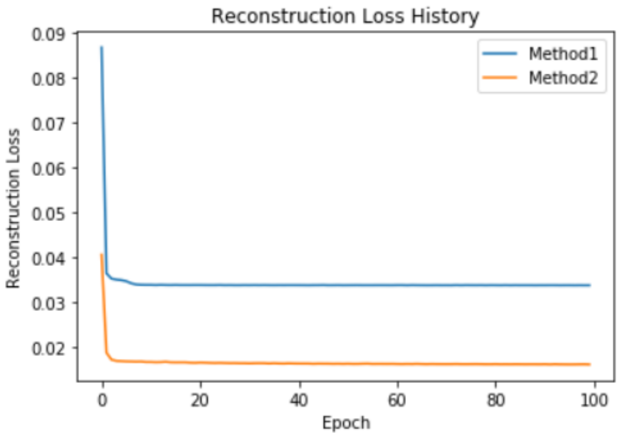
學號：R07725019 系級： 資管碩二 姓名：鄒雅雯

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。）

Autoencoder：

I. **3 Layer**：
nn.Conv2d(3, 10, 3, 2, 1), nn.Conv2d(10, 20, 3, 2, 1), nn.Conv2d(20, 5, 3, 2, 1)

II. **4 Layer**：
nn.Conv2d(3, 256, 3, 2, 1), nn.Conv2d(256, 128, 3, 2, 1),
nn.Conv2d(128, 80, 3, 2, 1), nn.Conv2d(80, 40, 3, 2, 1)



不論是 3 層還是 4 層的 autoencoder，很快就收斂了，但 4 層的 Reconstruction error 明顯較低。

降維方式 ：

I. PCA 18 維

II. TSNE 2 維

Result：以下 autoencoder 皆訓練 20 epoch

I. 3 Layer + PCA 18 維
Public score：0.73888 / Private score：0.73492

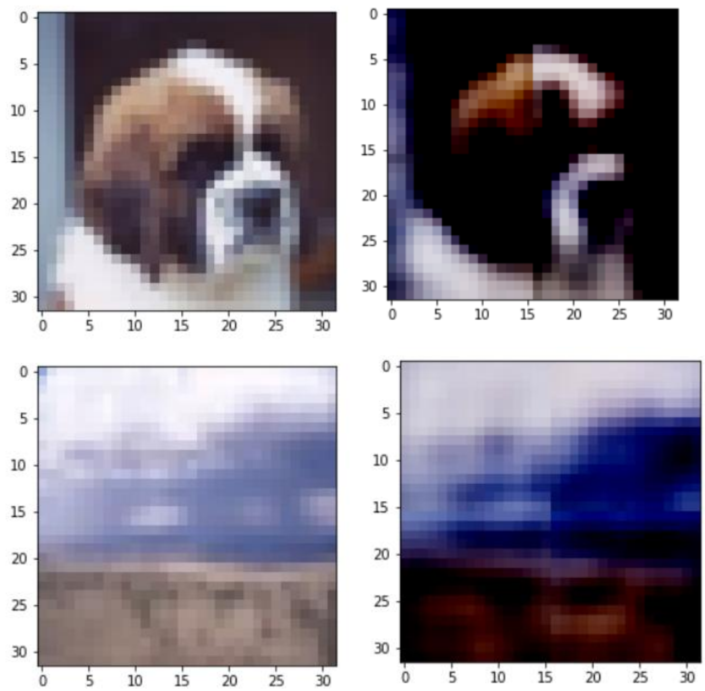
II. 3 Layer + TSNE 2 維
Public score：0.75185 / Private score：0.75238

III. 4 Layer + PCA 18 維
Public score：0.70222 / Private score：0.70317

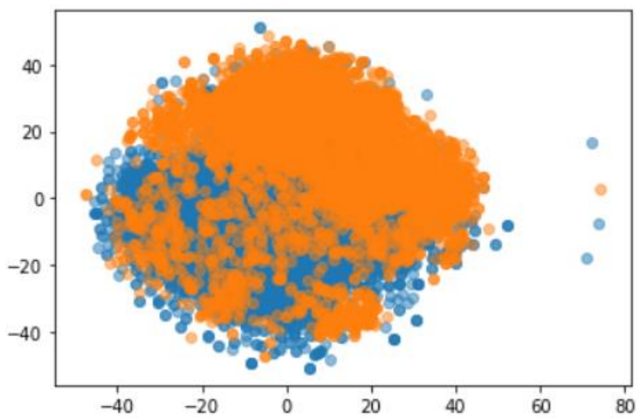
IV. 4 Layer + TSNE 2 維
Public score：0.79592 / Private score：0.80222

從結果中看出，TSNE 的表現不論搭配哪個 model，皆有較好的分數，在 4 Layer 的模型中尤其是，因為 4 Layer 模型產出的 feature 較 3 Layer 多，因此 PCA 只取 18 個 feature 明顯不足，至於 autoencoder 模型部分，從搭配 TSNE 降維的結果看來，4 層的模型表現較好。

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

*1

$$\begin{bmatrix} 1 & 4 & 3 & 1 & 5 & 7 & 9 & 3 & 11 & 10 \\ 2 & 8 & 12 & 8 & 14 & 4 & 8 & 8 & 5 & 11 \\ 3 & 5 & 9 & 5 & 2 & 1 & 9 & 1 & 6 & 7 \end{bmatrix} \quad \begin{aligned} u_1 &= 5.4 \\ u_2 &= 8 \\ u_3 &= 4.8 \end{aligned} \rightarrow \begin{bmatrix} -4.4 & -1.4 & -2.4 & -4.4 & -0.4 & 1.6 & 3.6 & -2.4 & 5.6 & 4.6 \\ -6 & 0 & 4 & 0 & 6 & -4 & 0 & 0 & -3 & 3 \\ -1.8 & 0.2 & 4.2 & 0.2 & -2.8 & -3.8 & 4.2 & -3.8 & 1.2 & 2.2 \end{bmatrix}$$

$$C = \frac{1}{9} XX^T = \begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix} \rightarrow \begin{vmatrix} 13.38-\lambda & 0.56 & 3.64 \\ 0.56 & 13.56-\lambda & 3.22 \\ 3.64 & 3.22 & 9.07-\lambda \end{vmatrix} = 0$$

$$\rightarrow \lambda = [\, 6.08 \;\; 12.92 \;\; 16.99 \,]$$

$$\begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 16.99 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.85 \\ -0.027 \\ -0.523 \end{bmatrix}$$

$$\begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 12.92 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.338 \\ 0.734 \\ -0.589 \end{bmatrix} \rightarrow \underline{\begin{bmatrix} -0.85 & 0.338 & 0.399 \\ -0.027 & 0.734 & -0.678 \\ -0.523 & -0.589 & -0.617 \end{bmatrix}} \#$$

$$\begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 6.08 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.399 \\ -0.678 \\ -0.617 \end{bmatrix}$$

$$X' = X^T \begin{bmatrix} -0.85 & 0.338 & 0.399 \\ -0.027 & 0.734 & -0.678 \\ -0.523 & -0.589 & -0.617 \end{bmatrix} = \underline{\begin{bmatrix} 4.85 & 1.088 & -0.26 & 3.64 & 1.64 & 0.73 & -5.26 & 4.03 & -5.32 & -5.15 \\ -4.83 & -0.59 & -0.35 & -1.6 & 5.92 & -0.16 & -1.26 & 1.43 & -1.02 & 2.46 \\ 3.42 & -0.68 & -6.26 & -1.88 & -2.5 & 5.69 & -1.15 & 1.38 & 3.53 & -1.55 \end{bmatrix}} \#$$

$$R_e = \begin{bmatrix} 4.85 & -4.83 \\ 1.088 & -0.59 \\ -0.26 & -0.35 \\ 3.64 & -1.6 \\ 1.64 & 5.92 \\ -0.73 & -0.16 \\ -5.26 & 4.03 \\ -5.32 & -1.02 \\ -5.15 & 2.46 \end{bmatrix} \begin{bmatrix} -0.85 & -0.027 & -0.523 \\ 0.338 & 0.734 & -0.589 \end{bmatrix} = \begin{bmatrix} -5.76 & -1.13 & 0.1 & -3.65 & 0.6 & -0.67 & 4.06 & -2.95 & 4.18 & 5.22 \\ -3.68 & 0.46 & -0.25 & -1.28 & 4.3 & -0.14 & -0.78 & 0.94 & -0.6 & 1.95 \\ 0.31 & -0.22 & 0.34 & -0.96 & -4.34 & -0.28 & 3.49 & -2.95 & 3.38 & 1.24 \end{bmatrix}$$

$$\text{reconstruction error} = \frac{\sqrt{\sum (X' - R_e)^2}}{10} = \underline{2.81} \#$$

*2

(a)

Let λ be a non-zero eigenvalue of AA^T and q be the eigenvector corresponding to λ. Then $(AA^T)q = \lambda q$.

Premultiplying both sides by A^T, $\Rightarrow A^T(AA^T)q = (A^TA)\cdot(A^Tq) = \lambda(A^Tq)$

Therefore, λ is an eigenvalue of A^TA with A^Tq as the eigenvector.

$\Rightarrow$ AA^T and A^TA have identical non-zero eigenvalues.

Let λ be an eigenvalue of AA^T and q be the eigenvector corresponding to λ.

$\rightarrow (AA^T)q = \lambda q \xrightarrow[q^T]{\text{同乘}} q^TAA^Tq = \lambda q^Tq \rightarrow \lambda = \dfrac{q^TAA^Tq}{q^Tq} = \dfrac{z^Tz}{q^Tq}$ $z = A^Tq$

$\because q^Tq > 0$ and $z^Tz \geq 0$ $\therefore \lambda \geq 0 \Rightarrow$ all eigenvalues of AA^T are non-negative

$x^TA^TAx = (Ax)^T(Ax) = \|Ax\|_2^2 \geq 0$, for all $x \in \mathbb{R}^n$

$y^TAA^Ty = (A^Ty)^T(A^Ty) = \|A^Ty\|_2^2 \geq 0$. for all $y \in \mathbb{R}^m$

$\Rightarrow$ A^TA and AA^T are symmetric positive semi definite matrices.

(b) data points : (1,2,3), (4,8,5), (3,12,9), (1,8,5), (5,14,2), (7,4,1), (9,8,9), (3,8,1), (11,5,6), (10,11,7) 如同 Q1

$\rightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \begin{bmatrix} 5.4 \\ 8 \\ 4.8 \end{bmatrix} \in \mathbb{R}^3$, $\Sigma = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T = \begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix}$ $\rightarrow$ symmetric

$\rightarrow$ eigenvalues = [6.08 12.92 16.99] by Q1 $\rightarrow$ non-negative eigenvalues

$\therefore$ Σ is a positive semi-definite matrix