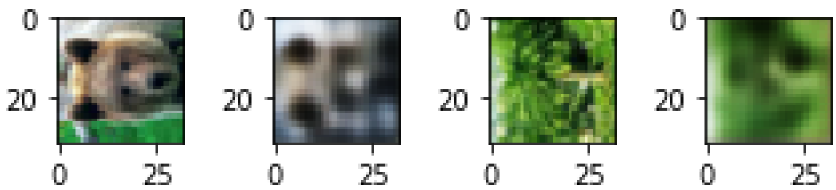
1. (1%) 請使用不同的 Autoencoder model，以及不同的降維方式(降到不同維度)，討論其 reconstruction loss & public / private accuracy。(因此模型需要兩種，降維方法也需要兩種，但 clustrering 不用兩種。)

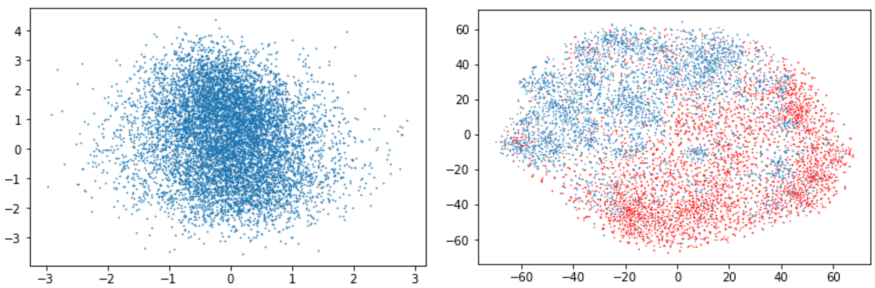
(1) latent 降到 32 維 使用 pca 降到 16 維 loss = 0.0473 accuracy = 0.7933/0.80518
(2) latent 降到 32 維 使用 t_sne 降到 2 維 loss = 0.0473 accuracy = 0.81444/0.81333
(3) latent 降到 64 維 使用 pca 降到 16 維 loss = 0.03 accuracy = 0.73825/0.74703
(4) latent 降到 64 維 使用 t_sne 降到 2 維 loss = 0,03 accuracu = 0.75623/0.76029

看起來在這筆資料下，兩種降維方式並不會差太多，而 loss 越低也不代表準確度就高

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

1.

Let $A = \begin{bmatrix} 1 & 4 & 3 & 1 & 5 & 7 & 9 & 3 & 11 & 10 \\ 2 & 8 & 12 & 8 & 14 & 4 & 8 & 8 & 5 & 11 \\ 3 & 5 & 9 & 5 & 2 & 1 & 9 & 1 & 6 & 7 \end{bmatrix}$ $\quad \bar{A} = \begin{bmatrix} 5.4 \\ 8 \\ 4.8 \end{bmatrix}$

① $A - \bar{A}$

② $\Sigma = \frac{1}{10}(\bar{A} - A)(A - \bar{A})^T = \begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix}$

(a) ③ eigen Value = 6.09 , $\underline{12.92}$, $\underline{16.99}$ eigen Vector = $\begin{bmatrix} -0.61 & -0.59 & -0.52 \\ -0.67 & 0.73 & -0.03 \end{bmatrix} = W$

(b) ④ $Z = W(A - \bar{A}) = \begin{bmatrix} -7.19 & 0.76 & -3.07 & 2.61 & -1.82 & 3.35 & -4.41 & 3.47 & -2.31 & -5.75 \\ 1.37 & -0.94 & -4.45 & -2.98 & -4.75 & 3.92 & 2.56 & -1.73 & 6.03 & 0.98 \end{bmatrix}$

⑤ $\|WW^T\| = 1 \Rightarrow (A - \bar{A})' = W^T Z$

⑥ error $= \frac{1}{10}[(A - \bar{A}) - (A - \bar{A})']^T[(A - \bar{A}) - (A - \bar{A})'] I$

(c) error = trace (error) $\approx$ 5.47

2.

(1) $A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ A_{m1} & \cdots & \cdots & A_{mn} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix}$, where $A_2 = [A_{21} \; A_{22} \; \cdots \; A_{2n}]_{1 \times n}$

$AA^T = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} [A_1^T \; A_2^T \cdots A_m^T] = \begin{bmatrix} A_1A_1^T & A_1A_2^T & \cdots & A_1A_m^T \\ A_2A_1^T & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ A_mA_1^T & \cdots & \cdots & A_mA_m^T \end{bmatrix}$

Since $A_1A_2^T = A_2A_1^T$, AA^T is a symmetric matrix.

Similarly , A^TA is symmetric.

Given $x \in \mathbb{R}^m$, $(x^TA)(A^Tx) \geq 0$ $\Rightarrow$ They are positive semi-definite.

Given $x \in \mathbb{R}^n$, $(x^TA^T)(Ax) \geq 0$

$AA^Tv = \lambda v \Rightarrow v^TAA^Tv = \lambda v^Tv \Rightarrow \lambda = \|A^Tv\|^2 / \|v\|^2 > 0$

$\Rightarrow$ They share same non-zero eigen-values.

(2) $X = \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ x_{12} & \ddots & \vdots \\ \vdots & & \vdots \\ x_{1m} & \cdots & x_{nm} \end{bmatrix}$ Claim: $\left(\frac{1}{n} X X^T = \Sigma \right)$

since Σ is a symmetric positive semi-definite matrix,

Σ can be rewite as GG^T by cholesky decomposition.

$\Rightarrow \exists \; X = n^{\frac{1}{2}} G \ni \frac{1}{n} X X^T = \frac{1}{n}(n^{\frac{1}{2}}G)(n^{\frac{1}{2}}G^T) = GG^T = \Sigma$ #.

(3) $L = tr(\Phi^T \Sigma \Phi) + \lambda(I_k - \Phi^T \Phi)$

$\frac{\partial L}{\partial \Phi} = (\Sigma + \Sigma^T)\Phi - 2\lambda\Phi = 0$

$\Rightarrow \Phi$ is the eigen-vector of $(\Sigma + \Sigma^T)$ #