學號：B05705006 系級： 資管四 姓名：李和維

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。）

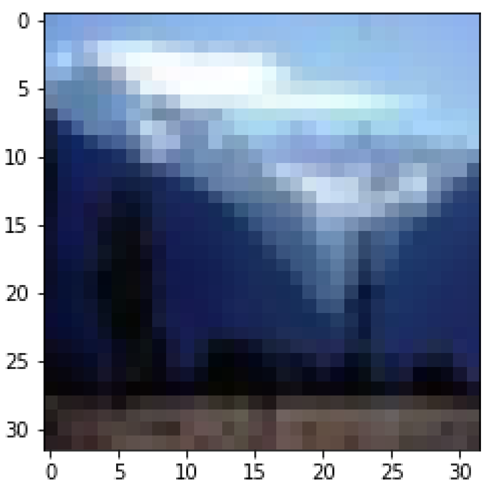
epoch 皆設為 200 或直到 reconstruction loss 收斂，lr = 3e-3

Autoencoder model	二次降維	reconstruction loss	public accuracy	private accuracy
Encoder: Conv2d(3 -> 32 -> 64) Linear(8*8*64 -> 50) Decoder: Linear(50 -> 8*8*64) Conv2d(64 -> 32 -> 3)	PCA (n=32)	0.026530	0.72000	0.71730
Same as above	TSNE (n=2)	0.026530	0.81296	0.81380
Encoder: Conv2d(3 -> 32 -> 64 -> 128 -> 256) Linear(2*2*256 -> 256 -> 128 -> 50) Decoder: Linear(50 -> 128 -> 256 -> 2*2*256) Conv2d(256 -> 128 -> 64 -> 32 -> 3)	PCA (n=32)	0.038232	0.70592	0.70809
Same as above	TSNE (=2)	0.038232	0.77703	0.77079

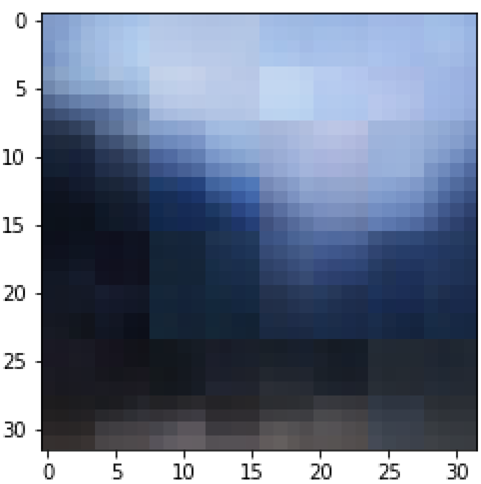
較多層的 Autoencoder 因為較難 train 所以在相同 epoch 與 lr 情況下表現不見得比較簡單的 Autoencoder 差；而二次降維使用 TSNE 的準確率普遍都比使用 PCA 來的佳。

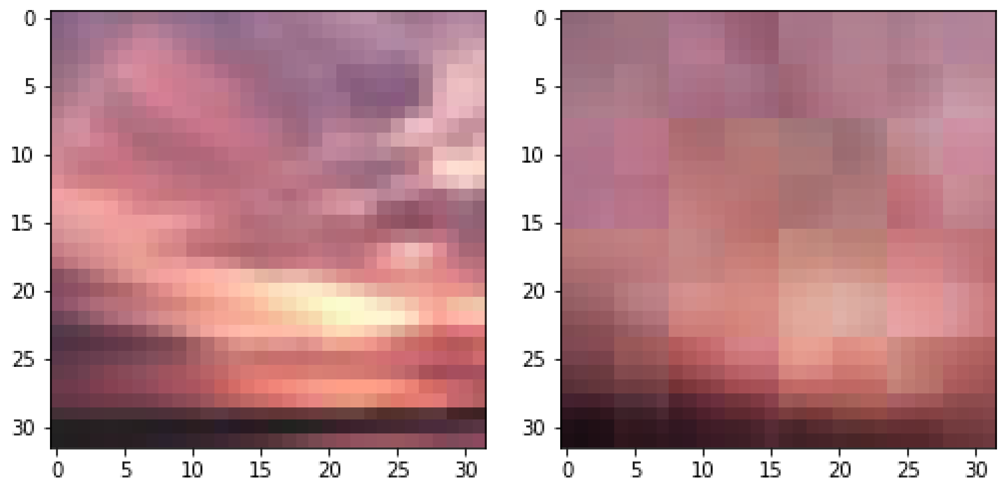
2. (1%) 從dataset選出2張圖，並貼上原圖以及經過autoencoder後reconstruct的圖片。

Original

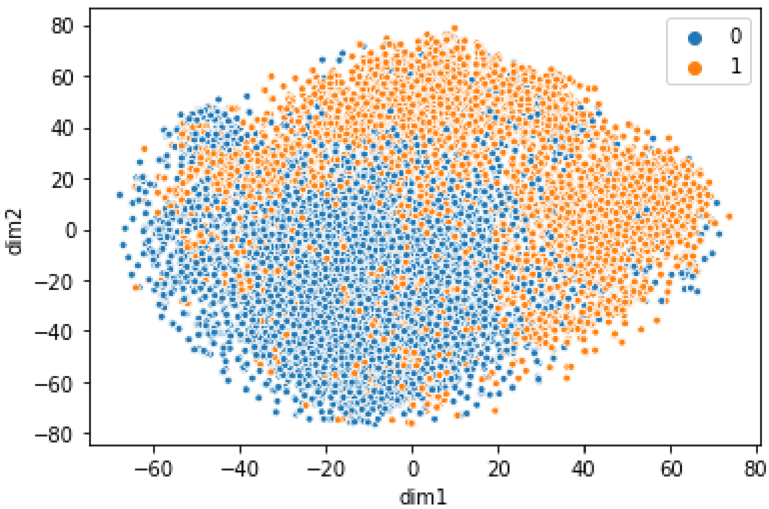


After reconstruction





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.

1.

$$\Sigma = \frac{1}{10} \sum (x_i - \bar{x})(x_i - \bar{x})^T = \begin{bmatrix} 12.04 & 0.5 & 3.28 \\ 0.5 & 12.2 & 2.9 \\ 3.28 & 2.9 & 8.16 \end{bmatrix}$$

$\lambda_1 = 15.3$, $u_1 = [-0.62, -0.59, -0.52]^T$,

$\lambda_2 = 11.63$, $u_2 = [-0.68, 0.73, -0.03]^T$,

$\lambda_3 = 5.47$, $u_3 = [0.4, 0.34, -0.85]^T$

(a) u_1, u_2, u_3 ※

(b) $z_i = W \times x_i$, $W = \begin{bmatrix} -0.62 & -0.59 & -0.52 \\ -0.68 & 0.73 & -0.03 \\ 0.4 & 0.34 & -0.85 \end{bmatrix}$

$z_1 = [-3.36, 0.71, 1.48]^T$, $z_2 = [-9.79, 3.03, -0.04]^T$,

$z_3 = [-13.62, 6.53, 2.42]^T$, $z_4 = [-7.94, 5.06, 1.16]^T$,

$z_5 = [-12.37, 6.84, -5.02]^T$, $z_6 = [-7.19, -1.84, -3.3]^T$,

$z_7 = [-14.96, -0.47, 1.37]^T$, $z_8 = [-7.05, 3.81, -3.05]^T$,

$z_9 = [-12.86, -3.95, -0.97]^T$, $z_{10} = [-16.3, 1.11, -1.75]^T$ ※

(c) $\frac{1}{10} \sum (x_i - \hat{y}_i)^2 = 6.064$ ※

2.

2.

(a) $(AA^T)^T = (A^T)^T A^T = AA^T \Rightarrow$ AA^T is symmetric

$(A^TA)^T = A^T(A^T)^T = A^TA \Rightarrow$ A^TA is symmetric

$z^T(AA^T)z = (A^Tz)^T A^T z = \|A^Tz\|^2 \geq 0$

$\Rightarrow$ AA^T is positive semi-definite

$z^T(A^TA)z = (Az)^T Az = \|Az\|^2 \geq 0$

$\Rightarrow$ A^TA is positive semi-definite

$(AA^T)x = \lambda x$, $\because \lambda \neq 0$, $A^TA(A^Tx) = \lambda(A^Tx)$, $\lambda = \|A\|^2$,

$(A^TA)x = \lambda x$, $\because \lambda \neq 0$, $AA^T(Ax) = \lambda(Ax)$, $\lambda = \|A\|^2$

No.
Date

(b) for all $x \in R^m$,

$Cov(X) = E[(X-\mu)(X-\mu)^T] = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T = \Sigma$,

let $M = (x-\mu)(x-\mu)^T \Rightarrow M$ is symmetric and positive semi-definite

$\therefore x = (x_1, \cdots x_n)$ 可為任意實數。

3.

No.
Date

3.

Find $g_{T+1}^{k}(x)$, $k=1\ldots K$, minimize $L(g_T^1, \cdots, g_T^k)$,

For every $k=1\cdots K$,

$$g_{T+1}^{k}(x) = g_T^k(x) - \eta \underline{\underline{\frac{\partial L(g)}{\partial g_T^k(x)}}},$$

$\rightarrow$ same direction

$$g_{T+1}^{k}(x) = g_T^k(x) + \underline{\underline{\alpha_{T+1}^{k} f_{T+1}(x)}}$$

$$\frac{\partial L}{\partial g_T^k(x)} = \sum_{i=1}^{n} \exp\left(\frac{1}{K-1}\sum_{k\neq \hat{y}_n} g_T^k(x_n) - g_T^{\hat{y}_n}(x_n)\right) \times D,$$

$$\left(D = \begin{cases} \frac{1}{K-1}, & \text{if } k\neq \hat{y}_n \\ -1, & \text{else} \end{cases}\right)$$

We want to find $f_{T+1}(x)$ maximizing

$$\sum_{i=1}^{n} \exp\left(\frac{1}{K-1}\sum_{k\neq \hat{y}_n} g_T^k(x_n) - g_T^{\hat{y}_n}(x_n)\right) \times D \times f_{T+1}(x)$$

We want to α_T^k minimizing $L(g)$,

Find α_T^k such that $\frac{\partial L}{\partial \alpha_T^k} = 0$ ※

$$* \frac{\partial L}{\partial \alpha_T^k} = \sum_{i=1}^{n} \exp\left(\frac{1}{K-1}\sum_{k\neq \hat{y}_n}\sum_{t=1}^{T}\alpha_t^k f_t(x_n) - \sum_{t=1}^{T}\alpha_t^{\hat{y}_n} f_t(x_n)\right) \times E.$$

$\left(E = f_t(x_n) \times \frac{1}{K-1} \text{ if } k \neq \hat{y}_n \text{ else } f_t(x_n)\right)$