學號：R07946013 系級： 資料科學碩二 姓名：吳泓毅

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。）
   (a) 將資料攤平以後用 linear layers + relu 做 autoencoder: 3072 – 1000 – 500 – 1000 -3072
   (b) 用 convolution 和 transpose conv :

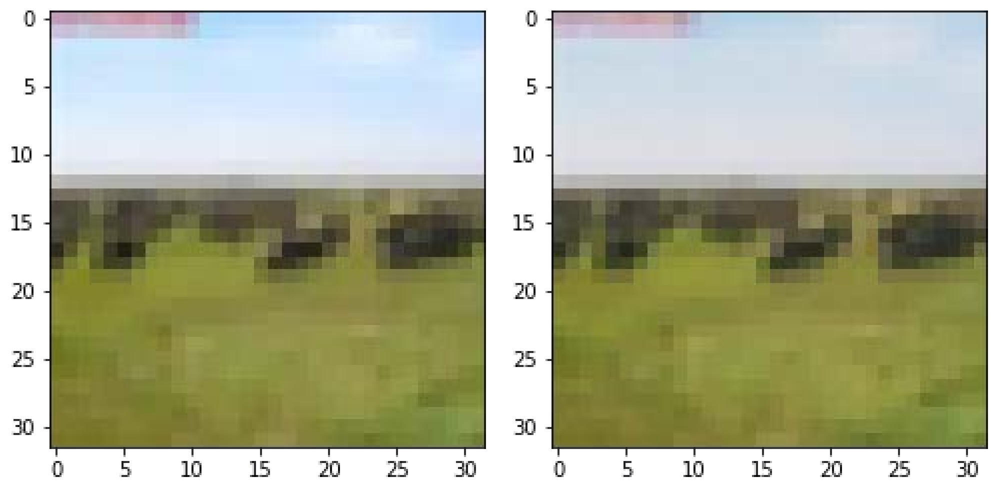
   ```
   nn.Conv2d(3,8, kernel_size=3, stride=2, padding=1)
   nn.Conv2d(8,24, kernel_size=3, stride=2, padding=1)
   nn.ConvTranspose2d(8, 3, kernel_size=2, stride=2)
   nn.ConvTranspose2d(8, 3, kernel_size=2, stride=2)
   ```

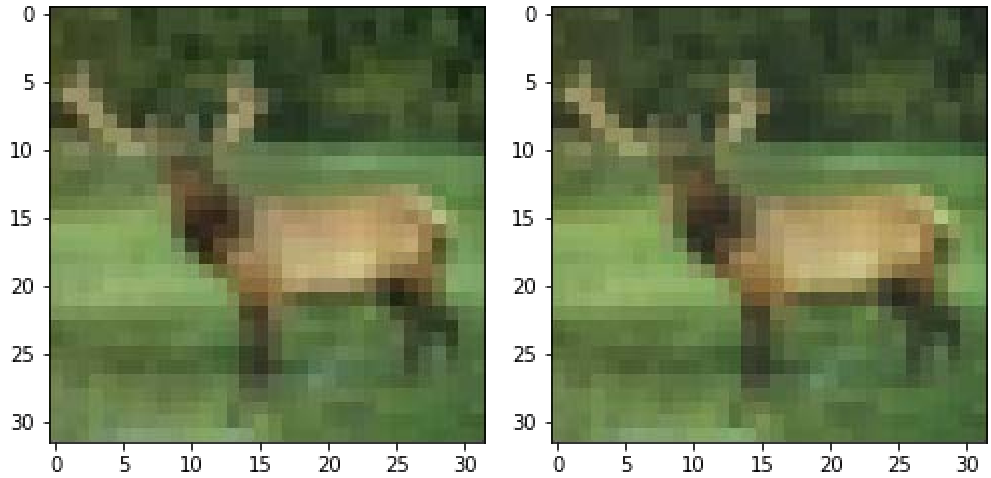
   (c) PCA
   (d) PCA + Tsne

	(a) Linear + relu (0.1204)	(b) Conv (loss=0.0335)
(c)PCA	0.55761/0.55814	0.63920/0.06300
(d)PCA + TSNE	0.77032/0.74556	0.77508/0.76852

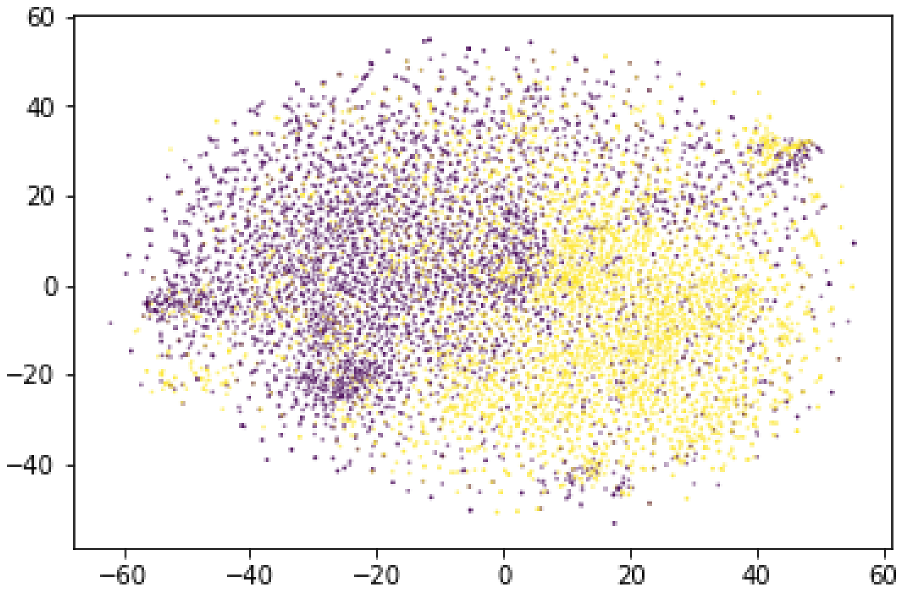
Clustering 的方式是用 Kmeans 和 ward 選高的。雖然 linear + relu 我只讓 loss 跑到 0.12 就停但實際上再繼續做下去雖然 loss 會降低但 clustering 的結果會變差。這邊也可以發現 loss 跟 clustering 的結果並沒有很一定的關係；而用兩層 cnn 的效果跟 linear 差不多但實際上 linear 用到的 parameter 數目遠高於 cnn 所以 cnn 還是比較的選擇。

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





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)
X =
array([[ 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]])

(a)
pca.components_ =
array([[-0.6165947 , -0.58881629, -0.52259579],
       [ 0.67817891, -0.73439013,  0.02728563],
       [ 0.39985541,  0.33758926, -0.85214385]])

(b)
pca.transform(X)
array([[ 7.18658682,  1.37323947, -2.25104047],
       [ 0.75871342, -0.94399334, -0.73022635],
       [-3.07034019, -4.45059025, -3.1883001 ],
       [ 2.60849751, -2.97853006, -1.92979259],
       [-1.82299166, -4.75401212,  4.25159619],
       [ 3.35457763,  3.91896138,  2.52755823],
       [-4.41464321,  2.55604371, -2.13952468],
       [ 3.46569126, -1.73131477,  2.27849363],
       [-2.31359638,  6.03371503,  0.2038499 ],
       [-5.75249521,  0.97648096,  0.97738622]])
(c)
reconstruct_X =
array([[ 1.90009072,  2.75992709,  1.08178971],
       [ 4.29198496,  8.24651657,  4.37774211],
       [ 4.27485905, 13.07633588,  6.28310968],
       [ 1.77163801,  8.65147726,  3.35553912],
       [ 3.29997625, 12.56470677,  5.62297154],
       [ 5.98934216,  3.14672348,  3.1538432 ],
       [ 9.85550052,  8.72228056,  7.17681721],
       [ 2.08893199,  7.23080501,  2.94160433],
       [10.91848951,  4.93118246,  6.17370944],
       [ 9.60918683, 10.6700449 ,  7.83287366]])

mean square error = 5.472032912651862
```

2. (a)

$A \in \mathbb{R}^{m\times n}$

symmetry : $(AB)^T = B^TA^T$, $\therefore \begin{cases} (AA^T)^T = (A^T)^TA^T = AA^T \\ (A^TA)^T = A^T(A^T)^T = A^TA \end{cases}$ $\Rightarrow$ Both are symmetric.

Positive semi-definite :

Let $v \in \mathbb{R}^m$, $v^TAA^Tv = \langle (A^Tv)^T, v^TA \rangle = \langle v^TA, v^TA \rangle = \| v^TA \| \geq 0$

$u \in \mathbb{R}^n$, $u^TA^TAu = \langle (Au)^T, u^TA^T \rangle = \langle u^TA^T, u^TA^T \rangle = \| u^TA^T \| \geq 0$

Share same non-zero eigenvalues :

Let $AA^Tx = \lambda x$, for some $\lambda \neq 0$, $x \in \mathbb{R}^m$. $x \neq 0$.

$A^TAA^Tx = A^TA(A^Tx) = \lambda(A^Tx)$ #.

(b) $\Sigma \in \mathbb{R}^{m\times m}$ and $\Sigma \geq 0$.

By Cholesky's decomposition, $\Sigma = LL^* = \sum_{i=1}^{k} g_i g_i^T$, where $k = \text{rank}(\Sigma)$ and g_i is the i-th column of L.

$\therefore \sum_{i=1}^{k} g_i g_i^T = \sum_{i=1}^{k} \left[(\tfrac{1}{\sqrt{2}} g_i)(\tfrac{1}{\sqrt{2}} g_i)^T + (-\tfrac{1}{\sqrt{2}} g_i)(-\tfrac{1}{\sqrt{2}} g_i)^T \right] = \Sigma$.

Also, $\sum_{i=1}^{k} \tfrac{1}{\sqrt{2}} g_i + (-\tfrac{1}{\sqrt{2}} g_i) = 0$

~~Denote $x_{i1} = \tfrac{1}{\sqrt{2}} g_i + \mu$, $x_{i2} = -\tfrac{1}{\sqrt{2}} g_i + \mu$~~

$\therefore$ Denote $g_i' = \begin{cases} \tfrac{1}{\sqrt{2}} g_i , & i = 1, \cdots, k \\ -\tfrac{1}{\sqrt{2}} g_i , & i = 1, \cdots, 2k. \end{cases}$

$\therefore \Sigma = \sum_{i=1}^{2k} g_i' g_i' = \frac{1}{n} \sum_{i=1}^{2k} (\tfrac{1}{\sqrt{n}} g_i')(\tfrac{1}{\sqrt{n}} g_i')^T$, $n = 2k$

$= \frac{1}{n} \sum_{i=1}^{2k} (\tfrac{1}{\sqrt{n}} g_i' + \mu - \mu)(\tfrac{1}{\sqrt{n}} g_i' + \mu - \mu)$, $n = 2k$

$\therefore \exists \left\{ \tfrac{1}{\sqrt{n}} g_i' + \mu \right\}_{i=1}^{2k}$ s.t. $\Sigma = \frac{1}{n} \sum_{i=1}^{2k} (\tfrac{1}{\sqrt{n}} g_i' + \mu - \mu)(\tfrac{1}{\sqrt{n}} g_i' + \mu - \mu)^T$.

Check: $\frac{1}{2k} \sum_{i=1}^{2k} (\tfrac{1}{\sqrt{n}} g_i' + \mu) = \frac{1}{2k} \sum_{i=1}^{k} \tfrac{1}{\sqrt{n}} (g_i - g_i) \cdot \tfrac{1}{\sqrt{2}} + \mu = \mu$ #.

2(c) $\begin{array}{ll}\min & tr(\Phi^T\Sigma\Phi) \\ s.t. & \Phi^T\Phi = I_m \\ & \Phi \in \mathbb{R}^{m\times k}\end{array} \Leftrightarrow \begin{array}{ll}\max & -tr(\Phi^T\Sigma\Phi) \\ s.t. & \Phi^T\Phi = I_m \\ & \Phi \in \mathbb{R}^{m\times k}\end{array} \Leftrightarrow \begin{array}{ll}\max & tr(\Phi^T(-\Sigma)\Phi) \\ s.t. & \Phi^T\Phi = I_m \\ & \Phi \in \mathbb{R}^{m\times k}\end{array}$

$\Leftrightarrow \begin{array}{ll}\max & tr(\Phi^T\Sigma'\Phi) - ck \\ s.t. & \Phi^T\Phi = I_k \\ & \Phi \in \mathbb{R}^{m\times k}\end{array}$, where $\Sigma' = cI - \Sigma'$ s.t. $\Sigma' \succeq 0$.

$\therefore tr(\Phi^T\Sigma'\Phi) = \frac{1}{n}\sum tr(\Phi^T x_i x_i^T \Phi) = \frac{1}{n}\sum \|\Phi^T x_i\|^2$. (By 2(b))

Known: Φ^T is the PCA direction.

$\therefore tr(\Phi^T\Sigma'\Phi) = \sum_{i=1}^{k} \lambda_i' = \sum_{i=1}^{k} (c - \lambda_i)$, where $c - \lambda_i$ is the i-th largest eigenvalue of Σ'.

$\therefore -\lambda_i$ is the i-th largest eigenvalue of $-\Sigma$.

$\therefore \begin{array}{ll} & tr(\Phi^T\Sigma\Phi) \\ s.t. & \Phi^T\Phi = I_k \\ & \Phi \in \mathbb{R}^{m\times k}\end{array} = \sum_{i=1}^{k} \lambda_i$, where λ_i is the i-th least eigenvalue of Σ.

(Known is by Eckart-Young-Mirsky Theorem)

3.

$$\frac{\partial L}{\partial g^s} = \sum_{i: \hat{y}_i \neq s} \exp\left[\frac{1}{K-1}\sum_{k \neq s} g_T^k(x_i) - g_T^s(x_i)\right] \cdot \left(-\frac{1}{K-1}\right) + \sum_{i: \hat{y}_i = s} \exp\left[\frac{1}{K-1}\sum_{\hat{y}_i \neq k} g_T^k(x_i) - g_T^{\hat{y}}(x_i)\right] \cdot \frac{1}{K-1}$$

$$= -\frac{1}{K-1} L_{(-s)} + \frac{1}{K-1} L_{(s)} = \frac{1}{K-1}\left(L_{(s)} - L_{(-s)}\right)$$

$$\therefore f_{T+1}^s(x) = -\frac{1}{K-1} L_{(-s)}(x) + \frac{1}{K-1} L_{(s)}(x) \Big|_{g(x) = g_T(x)}$$

$$\therefore g_{T+1}^s(x) = g_T^s + \alpha_T^s f_{T+1}(x)$$

$$\frac{\partial}{\partial \alpha_T^s} L(g_T^1, \ldots g_{T+1}^s, \ldots g_T^K) = \frac{\partial}{\partial \alpha_T^s} \sum_{i=1}^{n} \exp\left(\frac{1}{K-1} \sum_{\substack{\hat{y}_i = s \\ k \neq s}} g_T^k(x_i) - g_T^s(x) - \alpha_T^s f_T^s(x) + \sum_{\hat{y}_i \neq s} g_T^k(x_i) + \alpha_T^s f_T(x) - g_T^s(x)\right)$$

$$= \sum_{\substack{i \\ \hat{y}_i = s}} \exp\left(\frac{1}{K-1}\sum_{k \neq s} g_T^k(x_i) - g_T^s(x_i) - \alpha_T^s f_T^s(x)\right)\left[- f_T^s(x_i)\right] +$$

$$\sum_{\substack{i \\ \hat{y}_i \neq s}} \exp\left(\frac{1}{K-1}\left[\sum_{k \neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i) + \alpha_T^s f_T^s(x)\right]\right)\left[\frac{1}{K-1} f_T^s(x_i)\right]$$

$$= \sum_{\substack{i \\ \hat{y}_i = s}} \exp\left(\frac{1}{K-1}\sum_{k \neq s} g_T^k - g_T^s\right) \cdot \left(-f_T^s(x_i)\right) \exp\left(-\alpha_T^s f^s(x_i)\right) +$$

$$\sum_{\substack{i \\ \hat{y}_i \neq s}} \exp\left(\frac{1}{K-1}\left[\sum_{k \neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right]\right) \cdot \frac{1}{K-1} f_T^s(x_i) \cdot \exp\left(\alpha_T^s f_T^s(x_i)\right) = 0$$