Machine Learning - 2021Fall - HW5

系級：電機所碩一 學號：R10921098 姓名：楊仁傑

Kaggle Competetion

1. (1%) 請附上你在kaggle競賽上表現最好的降維以及分群方式，並條列五種不同降維維度的設定對應到的表現 (public / private accuracy)，auto-encoder 和 PCA 只要任一維度不一樣即可算是一種組合。

ANS：

在使用相同的AE架構，調整訓練的epoch以及PCA的維度(8092 -> n)，可以觀察出訓練300輪時，表現最好的是PCA n=1000，但當訓練變為500輪時，表現最好的反而是PCA n=500。

epoch	Loss
300	2.42
500	2.30

PCA	epoch	Public	Private
200	300	0.8396	0.8411
500	300	0.8331	0.8280
1000	300	0.8456	0.8424
200	500	0.8449	0.8462
500	500	0.8311	0.8244
1000	500	0.8422	0.8413

此外，另外嘗試調整learning rate去觀察AE收斂程度跟分類結果的差異性，對照上下兩個表可以得知，當AE的loss下降時，分類效果在固定參數的情況下並沒有比較好，甚至表現比較差。這也就呼應了助教在Kaggle中所提到的"好的autoencoder不代表你會cluster好兩類物品。因為model並不會知道你希望以什麼為依據做clustering。"

epoch	Loss
300	1.17
500	1.08

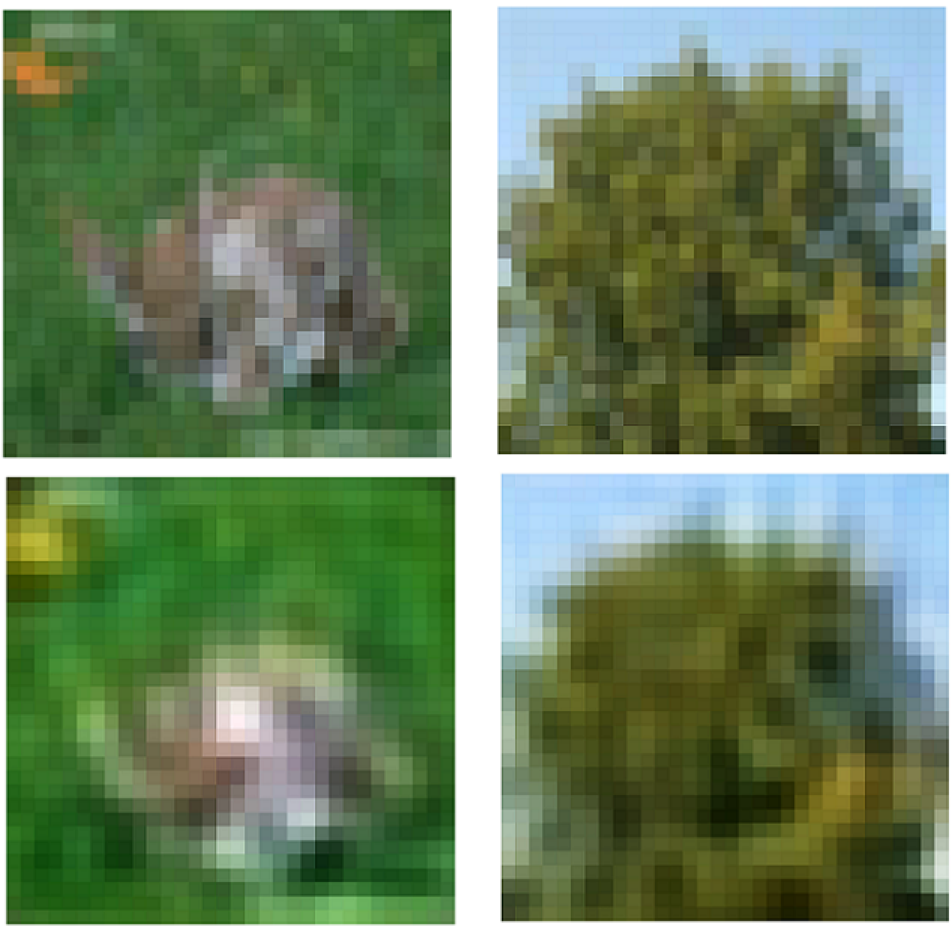
PCA	epoch	Public	Private
200	300	0.8067	0.8016
500	300	0.7671	0.7691
1000	300	0.8309	0.8253
200	500	0.8140	0.8207
500	500	0.8436	0.8396
1000	500	0.8327	0.8271

```
================================================================
            Conv2d-1           [-1, 64, 32, 32]           1,792
       BatchNorm2d-2           [-1, 64, 32, 32]             128
              ReLU-3           [-1, 64, 32, 32]               0
            Conv2d-4           [-1, 64, 32, 32]          36,928
       BatchNorm2d-5           [-1, 64, 32, 32]             128
              ReLU-6           [-1, 64, 32, 32]               0
            Conv2d-7          [-1, 128, 32, 32]          73,856
       BatchNorm2d-8          [-1, 128, 32, 32]             256
              ReLU-9          [-1, 128, 32, 32]               0
           Conv2d-10          [-1, 128, 32, 32]         147,584
      BatchNorm2d-11          [-1, 128, 32, 32]             256
             ReLU-12          [-1, 128, 32, 32]               0
        MaxPool2d-13          [-1, 128, 16, 16]               0
           Conv2d-14          [-1, 256, 16, 16]         295,168
      BatchNorm2d-15          [-1, 256, 16, 16]             512
             ReLU-16          [-1, 256, 16, 16]               0
           Conv2d-17          [-1, 256, 16, 16]         590,080
      BatchNorm2d-18          [-1, 256, 16, 16]             512
             ReLU-19          [-1, 256, 16, 16]               0
        MaxPool2d-20            [-1, 256, 8, 8]               0
           Conv2d-21            [-1, 512, 8, 8]       1,180,160
      BatchNorm2d-22            [-1, 512, 8, 8]           1,024
             ReLU-23            [-1, 512, 8, 8]               0
           Conv2d-24            [-1, 512, 8, 8]       2,359,808
      BatchNorm2d-25            [-1, 512, 8, 8]           1,024
             ReLU-26            [-1, 512, 8, 8]               0
        MaxPool2d-27            [-1, 512, 4, 4]               0
  ConvTranspose2d-28            [-1, 128, 8, 8]       1,638,528
             ReLU-29            [-1, 128, 8, 8]               0
  ConvTranspose2d-30           [-1, 64, 16, 16]         663,616
             ReLU-31           [-1, 64, 16, 16]               0
  ConvTranspose2d-32            [-1, 3, 32, 32]          55,491
             Tanh-33            [-1, 3, 32, 32]               0
================================================================
```

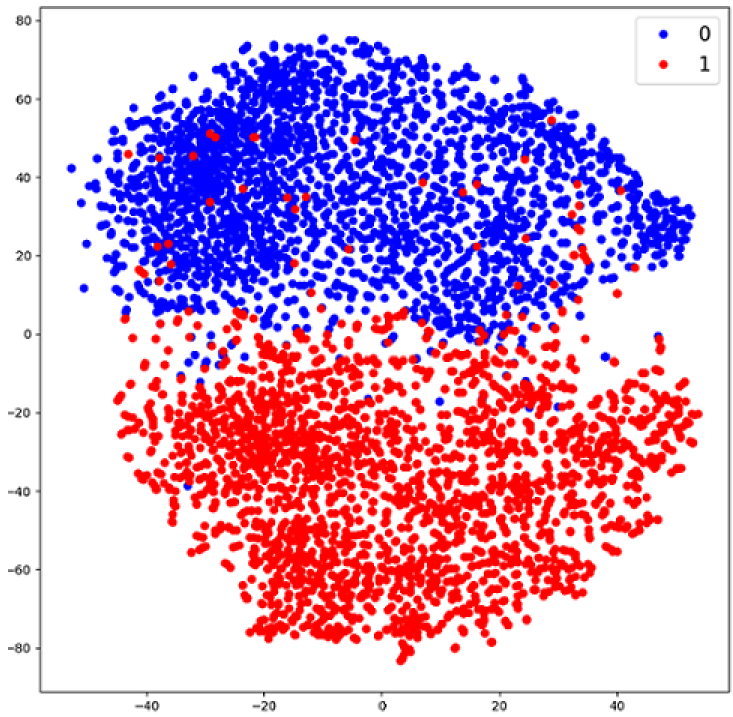
2. (1%) 從 kaggle 的 dataset 選出 2 張圖，並貼上原圖以及經過 autoencoder 後 reconstruct 的圖片；請將 visualization.npy 的檔案降維至二維平面並利用給定的 label 將資料上色 (前一半為 0；後半為 1)。

ANS：

[reconstruct]



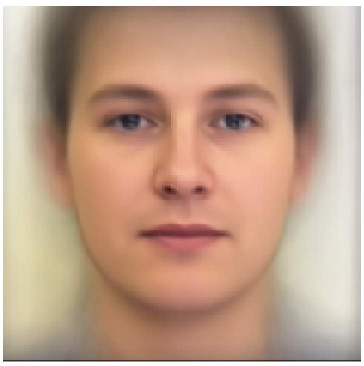
[visualization.npy]



Eigenface

3. (1%) 請畫出所有臉的平均以及 Eigenvalue 最大的前五個 Eigenfaces。

ANS：



4. (1%) 請從數據集中挑出任意五張圖片，並用上題前五大 Eigenfaces 進行 reconstruction，並畫出結果。

ANS：

Math Problem

5. (4%) Refer to math problem

[(https://hackmd.io/@g4HRMJCzQL2hzLedRcbVPQ/SyCBoc1qt)](https://hackmd.io/@g4HRMJCzQL2hzLedRcbVPQ/SyCBoc1qt)

ANS：

2. (a). $\mu = \frac{1}{10}\sum_{n=1}^{10} x_n = \begin{bmatrix} 5.4 \\ 8 \\ 4.8 \end{bmatrix}$

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

$\Rightarrow$ Eigenvalues, Eivenvector of S.

$\lambda_1 = 15.29$, $u_1 = \begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix}$

$\lambda_2 = 11.63$, $u_2 = \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix}$

$\lambda_3 = 5.47$, $u_1 = \begin{bmatrix} 0.40 \\ 0.34 \\ -0.85 \end{bmatrix}$

$\Rightarrow u_1, u_2, u_3$ are principal axes.

(b) $W = \begin{bmatrix} -0.02 & -0.59 & -0.52 \\ -0.68 & 0.73 & -0.03 \\ 0.4 & 0.34 & -0.85 \end{bmatrix}$

$Z_1 = \begin{bmatrix} -3.36 \\ 0.71 \\ 1.48 \end{bmatrix}$ $Z_2 = \begin{bmatrix} -9.78 \\ 3.03 \\ -0.04 \end{bmatrix}$ $Z_3 = \begin{bmatrix} -13.61 \\ 0.53 \\ 2.42 \end{bmatrix}$

$Z_4 = \begin{bmatrix} -7.94 \\ 5.06 \\ 1.16 \end{bmatrix}$ $Z_5 = \begin{bmatrix} -12.37 \\ 0.84 \\ -5.02 \end{bmatrix}$ $Z_6 = \begin{bmatrix} -7.19 \\ -1.84 \\ -3.3 \end{bmatrix}$ $Z_{10} = \begin{bmatrix} -10.30 \\ 1.11 \\ -1.75 \end{bmatrix}$

$Z_7 = \begin{bmatrix} -14.96 \\ -0.47 \\ 1.37 \end{bmatrix}$ $Z_8 = \begin{bmatrix} -7.08 \\ 3.81 \\ -3.05 \end{bmatrix}$ $Z_9 = \begin{bmatrix} -12.86 \\ -3.95 \\ -0.97 \end{bmatrix}$

(C) 3D → 2D

$$\frac{1}{10}\sum_{i=1}^{10}(x_i-\hat{y}_i)^2 = \underline{60.64} \ \#$$

3. (a). ① $\because (AA^T)^T = (A^T)^T A^T = AA^T$, $(A^TA)^T = A(A^T)^T = A^TA$

$\therefore AA^T$ and A^TA are both symmetric

② $\forall x \in R^m, x \neq 0$, $\forall y \in R^n, y \neq 0$.

$x^T(AA^T)x = (A^Tx)^T A^T x = ||A^Tx||^2 \geq 0.$

$x^T(A^TA)x = (Ax^T)Ax = ||Ax||^2 \geq 0.$

$\therefore A^TA$ and AA^T are both postive semi-definite

③ 令λ為AA^T的其中一組eigenvalue, $\lambda \neq 0$.

$\forall \underline{u} \in R^m, (AA^T)\underline{u} = \lambda \underline{u}$

$\Rightarrow (A^TA)(A^T\underline{u}) = A^T((AA^T)\underline{u})) = A^T(\lambda \underline{u}) = \lambda(A^T\underline{u})$

$\Rightarrow \lambda$同為(A^TA)的eigenvalue, $A^T\underline{u}$為其eigenvector

令Δ為AA^T的其中一組eigenvalue $\Delta \neq 0$

同理 $\forall \underline{u} \in R^n, (A^TA)\underline{u} = \Delta \underline{u}$

$\Rightarrow (AA^T)(A\underline{u}) = A((A^TA)\underline{u}) = A(\Delta \underline{u}) = \Delta(A\underline{u})$

$\Rightarrow \Delta$同為(AA^T)的eigenvalue, $A\underline{u}$為其eigenvector

$\therefore AA^T$ and A^TA有相同eigenvalue $\#$

Scanned with CamScanner

(b) $\because \Sigma$ is symmetric and positive definite.

$\Rightarrow \Sigma$ can be diagolize as

$\Sigma = U \Lambda U^T$, where $U \in R^{m\times m}$ is orthonormal

$\Lambda \in R^{m\times m}$ is diagonal

$\Lambda = \mathrm{diag}\left(\frac{\sigma_1^2}{n}, \frac{\sigma_2^2}{n}, \cdots \frac{\sigma_m^2}{n}\right)$ for $\begin{cases} \sigma_1 \cdots \sigma_m, \sigma_i \geq 0 \\ n, n > m \end{cases}$

$\Rightarrow \Sigma = U\left(\frac{1}{n} DD^T\right)U^T$ where $D \in R^{m\times n}$, $D = \begin{pmatrix} \sigma_1 & & & 0 & 0 \\ & \sigma_2 & & \vdots & \vdots \\ & & \ddots & 0 \cdots & 0 \\ 0 & & \sigma_m & 0 & 0 \end{pmatrix}$

$= \frac{1}{n} U D \underbrace{V^T V}_{I_n} D^T U^T$

$= \frac{1}{n} (UDV^T)(UDV^T)^T$

$= \frac{1}{n} X \cdot X^T$ (Let $X = UDV^T = (\hat{x}_1 \cdots \hat{x}_n)$)

$= \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i \hat{x}_i^T$

$= \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T$ ($\because$ Let $x_i = \hat{x}_i + \mu$)

There always exists V such that $\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i = 0$. , since:

$\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i = 0 \Leftrightarrow X\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} = 0 \Leftrightarrow UDV^T\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} = 0$

$\Leftrightarrow V^T\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} \in \mathrm{Null}(UD)$

since $\dim(\mathrm{Null}(UD)) = \mathrm{Nullity}(UD) = n - \mathrm{Rank}(UD) \geq n - m > 0$

$\Rightarrow \exists\, y = V^T\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}$

so, $\{x_1, x_2 \cdots x_n\}$ satisfies $\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i + \mu) + \mu$

$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T = \Sigma$

(c) since Σ and $\Phi\Phi^T$ are symmetric, exist orthogonal matrix.

Let $\phi = \underline{U_1}, \underline{U_2}, \ldots \underline{U_K}$ is a orthogonal vector set in R^m.

$(\phi\phi^T)\underline{U_i} = \phi(\phi^T \underline{U_i}) = \phi \cdot e_i = U_i = 1 \cdot \underline{U_i}$

eigenvalue = 0.

U_i is eigenvector of $\underline{\Phi\Phi^T}$ #