# ML HW4

學號：R08942025　系級：　電信碩一　姓名：徐瑋辰

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。）**

   AE models:

   A: DNN,  由 6 層 Fully Connected Layer 組成

   B: CNN,  由 Convolution/Deconvolution Layer 各 4 層組成

   降維方式:

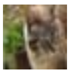
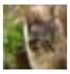
   C: PCA,  降至 256 維

   D: t-SNE,  降至 2 維

| | AC | AD |
|---|---|---|
| BC | Loss = 0.00073, Acc = 0.61630 | Loss = 0.00068, Acc = 0.73925 |
| BD | Loss = 0.00027, Acc = 0.68149 | Loss = 0.00031, Acc = 0.68259 |

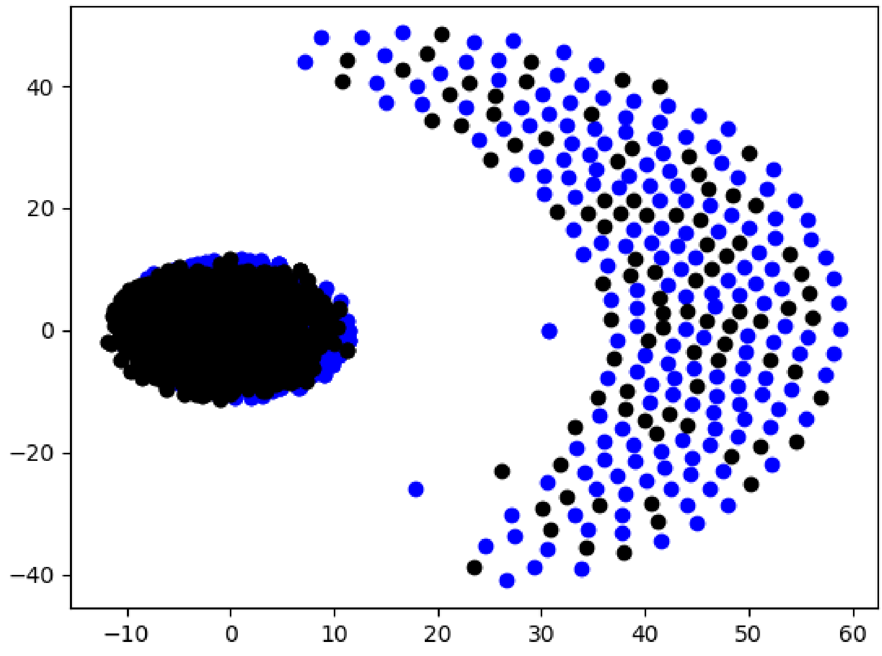
2. **從 dataset 選出 2 張圖，並貼上原圖以及經過 autoencoder 後 reconstruct 的圖片。**

   下方圖中  左圖為原圖  右圖為經過 autoencoder 後 reconstruct 的圖片

3. **請在二維平面上視覺化 label 的分佈。**

   經 AE, PCA, t-SNE 後之分布圖



4. **Refer to math problem**

1. (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) $\Bigg\}$ $\mu_x = 5.4$, $\mu_y = 8$, $\mu_z = 4.8$

(a) $$\Sigma = \frac{1}{10}\sum_{i=1}^{10}(x_i-\mu)(x_i-\mu)^T = \begin{bmatrix} 12.04 & 0.50 & 3.28 \\ 0.50 & 12.20 & 2.90 \\ 3.28 & 2.9 & 8.16 \end{bmatrix}$$

$$\Sigma = U \Lambda U^T \Rightarrow U = \begin{bmatrix} 0.6166 & -0.6782 & -0.3999 \\ 0.5888 & 0.7344 & -0.3376 \\ 0.5226 & -0.0273 & 0.8521 \end{bmatrix}$$

$\Rightarrow u_1 = (0.6166,\ 0.5888,\ 0.5226)$

$u_2 = (-0.6782,\ 0.7344\ \ -0.0273)$

$u_3 = (-0.3999,\ -0.3376,\ 0.8521)$

1 (b) samples $\to$ principle components $(= u_i^T x)$

$u_1 = (0.6166,\ 0.5888,\ 0.5226)$

$u_2 = (-0.6782,\ 0.7344\ \ -0.0273)$

$u_3 = (-0.3999,\ -0.3376,\ 0.8521)$

| | | | | |
|---|---|---|---|---|
| (1,2,3) | $\to$ | -1.9393 | 1.0448 | 3.0245 |
| (4,8,5) | $\to$ | -4.9583 | 6.5424 | 6.1328 |
| (3,12,9) | $\to$ | -9.8871 | 7.5408 | 8.9097 |
| (1,8,5) | $\to$ | -6.8081 | 4.7760 | 4.5650 |
| (5,14,2) | $\to$ | -7.2112 | 12.5504 | 3.9353 |
| (7,4,1) | $\to$ | 1.2036 | 6.7217 | 4.4012 |
| (9,8,9) | $\to$ | -3.4148 | 8.1362 | 12.1544 |
| (3,8,1) | $\to$ | -3.9755 | 7.3040 | 2.2016 |
| (11,5,6) | $\to$ | 0.9925 | 8.1234 | 10.7250 |
| (10,11,7) | $\to$ | -4.0930 | 11.6033 | 10.8908 |

1 (c) reconstructed error $= 0.1894 \times 10^{-27}$

2. (a) $A \in R^{m\times n}$ by $(AB)^T = B^T A^T$,

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

$(AA^T)^T = (A^T)^T A^T = AA^T \rightarrow$ symmetric

Let $\lambda$ be a non-zero eigenvalue of $AA^T$, $q$ be eigenvector of $\lambda$

$\Rightarrow (AA^T)\cdot q = \lambda q$

$\Rightarrow A^T(AA^T)q = (A^TA)(A^Tq) = \lambda(A^Tq)$

$\Rightarrow \lambda$ is an eigenvalue of $A^TA$, $A^Tq$ is the eigenvector

$(AA^T)q = \lambda q$

$\rightarrow q^T(AA^T)q = \lambda q^Tq$

$\rightarrow \lambda = \dfrac{q^T(AA^T)q}{q^Tq} = \dfrac{z^Tz}{q^Tq}$ $(z = A^Tq)$

$\Rightarrow \because q^Tq > 0$ and $z^Tz \geq 0$

$\Rightarrow \lambda \geq 0$, $AA^T$ is positive semi-definite

同理, $A^TA$ is also positive semi-definite

2 (b) Let $X = [x_1 \ x_2 \ \dots \ x_n] \in R^{m\times n}$

If $X$ 可表示为 $UDV^T$, $U \in R^{m\times m}$, $D \in R^{m\times n}$, $V \in R^{n\times n}$

$U, V$ are orthogonal. $D$ is diagonal

$$\Rightarrow \hat{X} = [\sum_{i=1}^{n} u_iu_i^T x_1 \quad \sum_{i=1}^{n} u_iu_i^T x_2 \quad \dots \quad \sum_{i=1}^{n} u_iu_i^T x_n]$$

$$= \sum_{i=1}^{n} u_iu_i^T [x_1 \ x_2 \ \dots \ x_n] = \sum_{i=1}^{n} u_iu_i^T X$$

$$= \sum_{i=1}^{n} u_iu_i^T UDV^T$$

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

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2 (C) Let $X = [x_1 \ x_2 \ \dots \ x_N]$

$\Sigma = \frac{1}{N} X X^T = U \Lambda U^T$

$\mathrm{Trace}(\Phi^T \Sigma \Phi)$

$= \frac{1}{N} \mathrm{Trace}(\Phi^T X X^T \Phi)$

$= \frac{1}{N} \| \Phi^T X \|_F^2$

$= \frac{1}{N} \sum_{i=1}^{N} \| \Phi^T x_i \|_F^2$

$= \frac{1}{N} \sum_{i=1}^{N} \| \hat{x}_i^{(S)} \|^2$

$\geq \frac{1}{N} \sum_{i=1}^{N} \| \hat{x}_i^{(PCA)} \|^2$

$\Phi_{opt} = [\hat{u}_1 \ \dots \ \hat{u}_k]$

proof:

$\mathrm{Trace}(\Phi \Sigma \Phi) = \mathrm{Trace}(\Phi^T U \Lambda U^T \Phi)$

(let $B = U^T \Phi$) $= \mathrm{Trace}(B^T \Lambda B)$

Let $\beta_j = \sum_{i=1}^{k} (B_{ji})^2$

$B^T B = \Phi^T U U^T \Phi = I_k$

$\Rightarrow \|\beta\|_1 = \sum_{j=1}^{m} \beta_j = \mathrm{Tr}(B B^T) = \mathrm{Tr}(I_k) = k$

$\Rightarrow \mathrm{Trace}(\Phi^T \Sigma \Phi) \geq \min_{\substack{\beta \in [0,1]^m \\ \|\beta\|_1 \geq k}} \sum_{j=1}^{m} \lambda_j \beta_j = \sum_{j=1}^{k} \lambda_j$

Set $\Phi_{opt} = [\hat{u}_1 \dots \hat{u}_k]$ where $\Sigma \hat{u}_i = \lambda_i \hat{u}_i$

$\Rightarrow \mathrm{Trace}(\Phi_{opt}^T \Sigma \Phi_{opt}) = \sum_{j=1}^{k} \lambda_j$ ※.

3. Algorithm

$g_0 = 0$

For $t = 0, 1, \dots, \frac{T}{k} - 1$

For $k = 1, 2, \dots, K$

$t = l \cdot K + k$

Find $f_t \in F$ s.t.

$\frac{\partial}{\partial \alpha} L(g_{t-1}^1, g_{t-1}^2, \dots, g_{t-1}^{k-1}, g_{t-1}^k + \alpha f_t, g_{t-1}^{k+1}, \dots, g_{t-1}^K)$ is minimized

Find $\alpha$ s.t.

$L(g_{t-1}^1, \dots, g_{t-1}^{k-1}, g_{t-1}^k + \alpha f_t, g_{t-1}^{k+1}, \dots, g_{t-1}^K)$ is minimized.

$L(g_T^1 \dots g_T^k) = \sum_{i=1}^{n} \exp\left( \frac{1}{K-1} \sum_{k \neq y_i} g_T^k(x_i) - g_T^{y_i}(x_i) \right)$

$g_T^k(x) = \sum_{t=1}^{T} a_t^k f_t(x)$

$h(x) = \underset{k}{\mathrm{argmax}} \ g_T^k(x)$