ML 2019 Hw4 Report

學號：b05705042 系級：資管四 姓名：皇甫立翔

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。

(1) Autoencoder (CNN : $3 \rightarrow 8 \rightarrow 16 \rightarrow 32$, Linear : $512 \rightarrow 256 \rightarrow 128$) + tSNE(2維)

reconstruction loss :
public accuracy : 0.82703
private accuracy : 0.82126

(2) Autoencoder (CNN : $3 \rightarrow 8 \rightarrow 16 \rightarrow 32 \rightarrow 64$, Linear : $512 \rightarrow 128$) + tSNE(2維)

reconstruction loss : 0.09422
public accuracy : 0.78238
private accuracy : 0.79333

(3) Autoencoder (CNN : $3 \rightarrow 8 \rightarrow 16 \rightarrow 32$, Linear : $512 \rightarrow 256 \rightarrow 128$) + PCA(32維)

reconstruction loss : 0.12726
public accuracy : 0.51222
private accuracy : 0.52174

(4) Autoencoder (CNN : $3 \rightarrow 8 \rightarrow 16 \rightarrow 32 \rightarrow 64$, Linear : $512 \rightarrow 128$) + PCA(32維)

reconstruction loss : 0.09422
public accuracy : 0.52814
private accuracy : 0.51968

從結果來看，tSNE 表現跟 PCA 差了不只一個次元，然後 reconstruction loss 越小不一定越好。

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

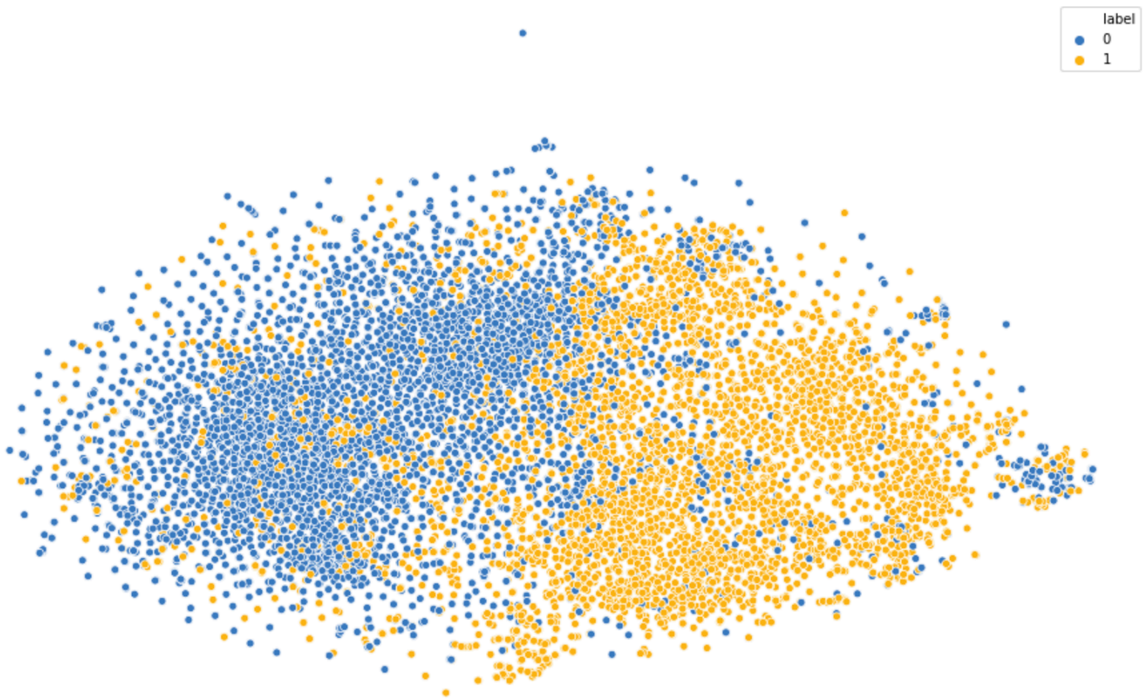
圖片一(原圖)

圖片一(reconstruct)

圖片二(原圖)

圖片二(reconstruct)

3. 請在二維平面上視覺化 label 的分佈。



(4. math problem 在下面。)

4. Refer to math problem。

1. $\text{cov}(X) = \frac{1}{10}\sum_{i=1}^{10}(x_i-\mu)(x_i-\mu)^T$. $\mu = [5.4 \;\; 8 \;\; 4.8]$

$$= \begin{bmatrix} 12.04 & 0.5 & 7.28 \\ 0.5 & 12.2 & 2.9 \\ 7.28 & 2.9 & 8.16 \end{bmatrix}$$

→ eigenvalues and eigenvectors of cov(X)

$\lambda_3 = 5.47$ $\quad \lambda_2 = 11.63$ $\quad \lambda_1 = 15.3$

$\mu_3 = \begin{bmatrix} 0.4 \\ 0.34 \\ -0.85 \end{bmatrix}$ $\quad \mu_2 = \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix}$ $\quad \mu_1 = \begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix}$

(a) The principal axes =

$\mu_1 = \begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix}$ $\quad \mu_2 = \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix}$ $\quad \mu_3 = \begin{bmatrix} 0.4 \\ 0.34 \\ -0.85 \end{bmatrix}$ *

(b)

$W = \begin{bmatrix} -0.62 & -0.59 & -0.52 \\ -0.68 & 0.73 & -0.03 \\ 0.4 & 0.34 & -0.85 \end{bmatrix}$, $z_i = W \cdot X$ (i^{th} principal component)

$z_1 = \begin{bmatrix} -3.36 \\ 0.71 \\ 1.48 \end{bmatrix}$ $\quad z_4 = \begin{bmatrix} -7.94 \\ 5.06 \\ 1.16 \end{bmatrix}$ $\quad z_7 = \begin{bmatrix} -14.96 \\ -0.47 \\ 1.37 \end{bmatrix}$ $\quad z_{10} = \begin{bmatrix} -16.3 \\ 1.11 \\ -1.75 \end{bmatrix}$

$z_2 = \begin{bmatrix} -9.79 \\ 3.03 \\ -0.04 \end{bmatrix}$ $\quad z_5 = \begin{bmatrix} -12.37 \\ 6.84 \\ -5.02 \end{bmatrix}$ $\quad z_8 = \begin{bmatrix} -7.08 \\ 3.81 \\ -3.05 \end{bmatrix}$

$z_3 = \begin{bmatrix} -13.62 \\ 6.53 \\ 2.42 \end{bmatrix}$ $\quad z_6 = \begin{bmatrix} -7.19 \\ -1.84 \\ -3.3 \end{bmatrix}$ $\quad z_9 = \begin{bmatrix} -12.86 \\ -3.95 \\ -0.97 \end{bmatrix}$ *

(c)

average reconstruction error

$$= \frac{1}{10}\sum_{i=1}^{10}(x_i - y_i)^2$$

$= 6.064$ *

2、

(a) $A \in R^{m\times n}$

(1) symmetric：

① for AA^T : $(AA^T)^T = (A^T)^T A^T = AA^T$, by definition of symmetry, AA^T is symmetric

② for A^TA : $(A^TA)^T = A^T(A^T)^T = A^TA$, by definition of symmetry, A^TA is symmetric ✱

(2) positive semi-definite：

z 為所有非零實係數向量, $\in R^m$ or R^n

① for AA^T : $z^T AA^T z = \underbrace{(A^Tz)^T A^Tz}_{\text{square of the inner product of } A^T, z\text{, thus must } \geq 0.} = \|A^Tz\|^2 \geq 0$

② for A^TA : $z^T A^TAz = \underbrace{(Az)^T Az}_{\text{square of inner product of } A, z\text{, thus must } \geq 0.} = \|Az\|^2 \geq 0$

⇒ AA^T and A^TA are both positive semi-definite matrix. ✱

(3) share same non-zero eigenvalue：

① for AA^T : $AA^Tx = \lambda x$, if $\lambda \neq 0$, $A^TA(A^Tx) = \lambda(A^Tx)$. $\lambda = A^TA = \|A\|^2$

② for A^TA : $A^TAx = \lambda x$, if $\lambda \neq 0$, $AA^T(Ax) = \lambda(Ax)$, $\lambda = AA^T = \|A\|^2$

⇒ AA^T and A^TA share the same non-zero eigenvalue. ✱

(b)

Let $x = (x_1, x_2 \cdots x_n) \in R^m$

By definition, covariance matrix of $x = E[(x-\mu)(x-\mu)^T]$. 其中 $\mu = E[x] = \frac{1}{n}\sum_{i=1}^{n} x_i$

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

⇒ let $M = (x-\mu)(x-\mu)^T$ and M is thus a symmetric positive semi-definite matrix.

for all 非零實係數向量 v. $v^TMv \geq 0$.

⇒ $v^TMv \geq 0$

⇒ $E[v^TMv] \geq 0$

⇒ $v^TE[M]v \geq 0$, 其中 $E[M] = E[(x-\mu)(x-\mu)^T] = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T = \Sigma$

⇒ $v^T\Sigma v \geq 0$

⇒ 由此可知, Σ is a symmetric positive semi-definite matrix 得證. ✱

(c)

Let $X = (x_1, x_2 \cdots x_n)$

s.t. $\Sigma = \frac{1}{N}XX^T = U\Lambda U^T$, then

$\mathrm{Trace}(\phi^T\Sigma\phi) = \frac{1}{N}\mathrm{Trace}(\phi^TXX^T\phi)$

$= \frac{1}{N}\|\phi^TX\|_F^2$

$= \frac{1}{N}\sum_{i=1}^{n}\|\phi^Tx_i\|^2$

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

⇒ $0 \leq \frac{1}{N}\sum_{i=1}^{n}\|\hat{x}_i^{(S)}\|^2 \leq \frac{1}{N}\sum_{i=1}^{n}\|\hat{x}_i^{(PCA)}\|^2$

3. objective: find $g_{T+1}^{k}(x)$. $\forall k \in 1\sim K$,

Minimize $Loss(g_T^1, g_T^2, \cdots g_T^k)$

$$\begin{cases} g_{T+1}^{k}(x) = g_T^k(x) - \eta \dfrac{\partial L(g)}{\partial g_T^k(x)} \\ g_{T+1}^{k}(x) = g_T^k(x) + \alpha_{T+1}^{k} f_{T+1}(x) \end{cases}$$

same direction

其中 $\dfrac{\partial L(g)}{\partial g_T^k(x)} = \sum_{i=1}^{n} \exp\left(\frac{1}{K-1} \sum_{k \neq y_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right) \cdot D$, $D = \begin{cases} \text{if } k \neq y_i, \frac{1}{K-1} \\ \text{else}, -1 \end{cases}$

$\Rightarrow$ then find $f_{t+1}(x)$ maximizing $\sum_{i=1}^{n} \exp\left(\frac{1}{K-1} \sum_{k \neq y_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right) \cdot D \cdot f_{t+1}(x)$

and find α_{T+1}^{k} minimizing $L(g)$ so that $\dfrac{\partial L}{\partial \alpha_T^k} = 0$.

其中 $\dfrac{\partial L}{\partial \alpha_T^k} = \sum_{i=1}^{n} \exp\left(\frac{1}{K-1} \sum_{k \neq y_i} \sum_{t=1}^{T} \alpha_t^k f_t(x_i) - \sum_{t=1}^{T} \alpha_t^{\hat{y}_i} f_t(x_i)\right) \cdot E$, $E = \begin{cases} \text{if } k \neq y_i, \frac{f_t(x_i)}{K-1} \\ \text{else}, f_t(x_i) \end{cases}$