EEML HWV

B06702064 會計五 林聖硯

最佳模型網址: https://drive.google.com/file/d/1h1LVKBVM7lZiSlxTaNmCuQDQsNEH_lV8/view?usp=sharing (https://drive.google.com/file/d/1h1LVKBVM7lZiSlxTaNmCuQDQsNEH_lV8/view?usp=sharing)

(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。

auto-encoder架構	分群方式	public acc	private acc
4 layers VAE (embed dim = 256)	tsne降成兩維->K-means	82.644%	82.444%
4 layers VAE (embed dim = 256)	PCA降成50維 ->tsne降成兩維->K-means	82.911%	83.066%
3 layers AE (embed dim = 256)	tsne降成兩維->K-means	71.955%	70.911%
3 layers AE (embed dim = 256)	PCA降成50維 ->tsne降成兩維->K-means	81.555%	81.266%
4 layers AE (embed dim = 128)	PCA降成45維 ->tsne降成兩維->K-means	83.733%	83.333%
4 layers AE (embed dim = 128)	PCA降成50維 ->tsne降成兩維->K-means	82.644%	82.444%
4 layers AE (embed dim = 128)	PCA降成65維 ->tsne降成兩維->K-means	79.333%	79.755%

觀察

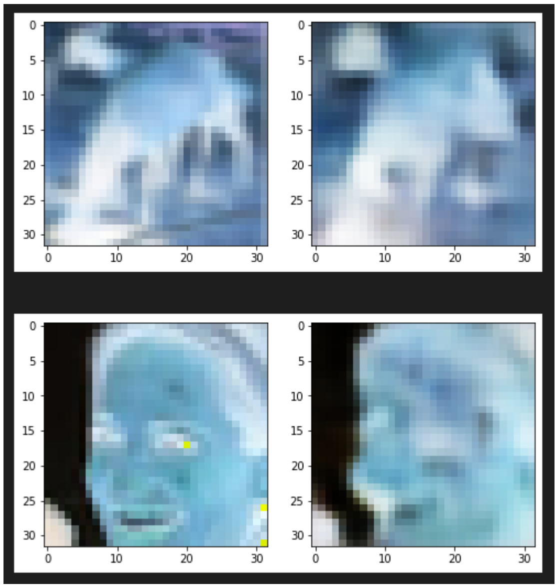
- t-sne的特性能有效將兩類的圖片拉開
- PCA降維後能使t-sne的表現再更進一步

- latent dimension 128維之表現比256維更好，推測一張圖片不需要過多的維度就能夠被extract得很好

(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)。

左圖: 原圖

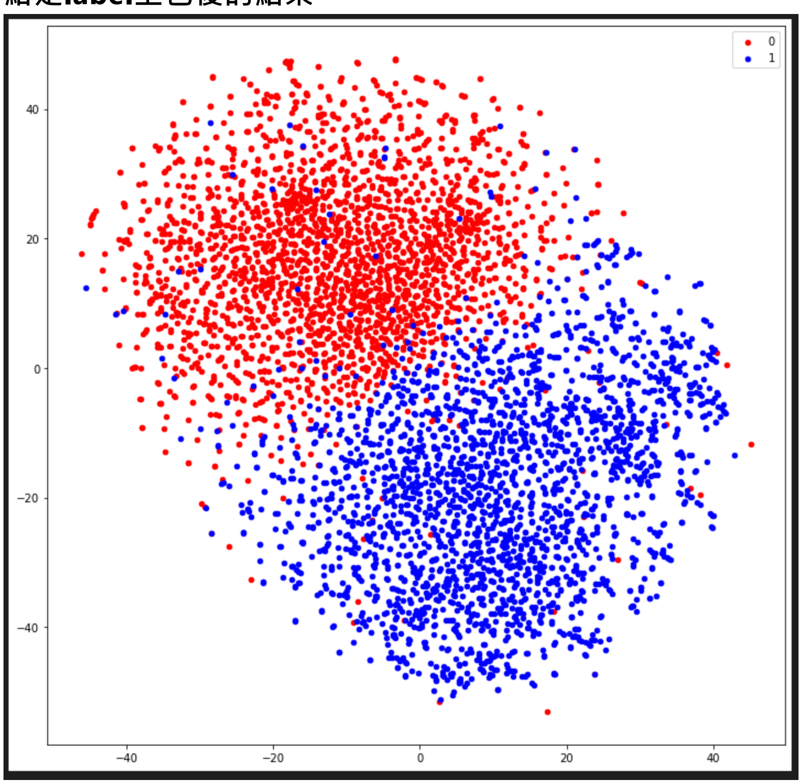
右圖: 經過AE後reconstruct的圖片



降維步驟: AE -> PCA(dim=45) -> t-sne(dim=2)

使用最好的model結果預測validation set，準確度高達95.78%

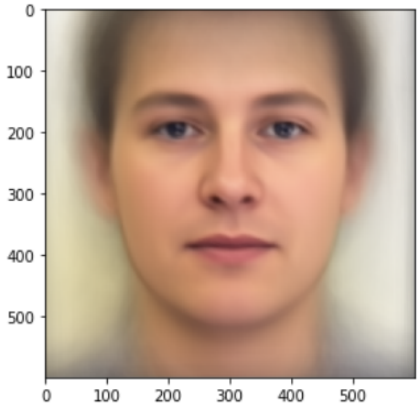
給定label上色後的結果



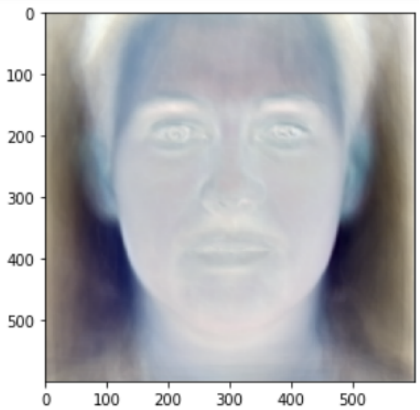
Eigenface

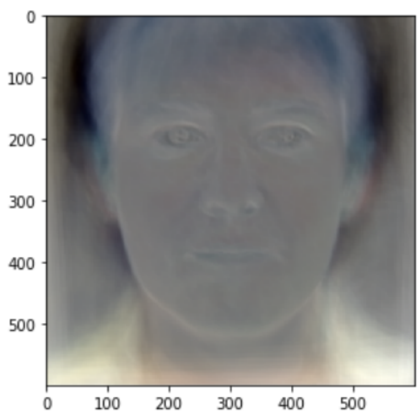
(1%) 請畫出所有臉的平均以及 Eigenvalue 最大的前五個 Eigenfaces。

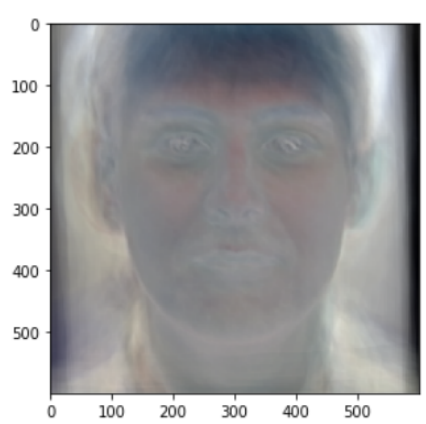
所有臉的平均

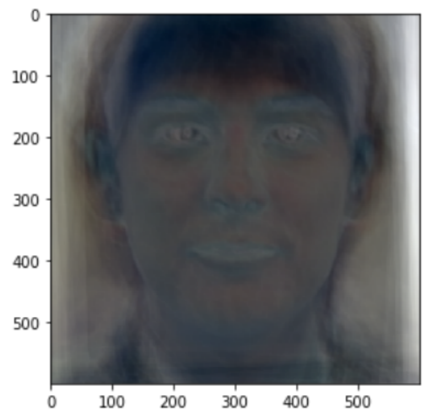


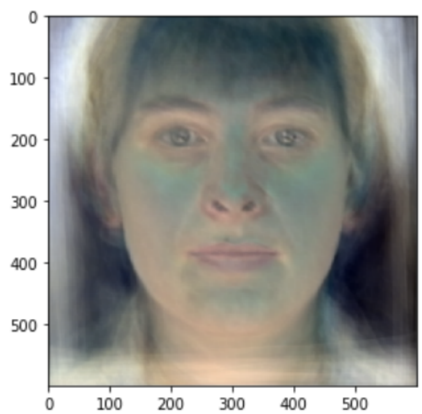
最大的前五個Eigenfaces





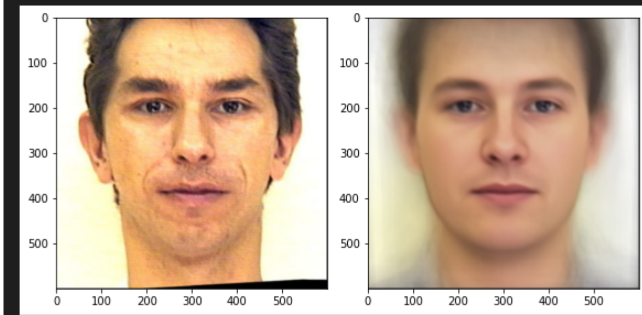




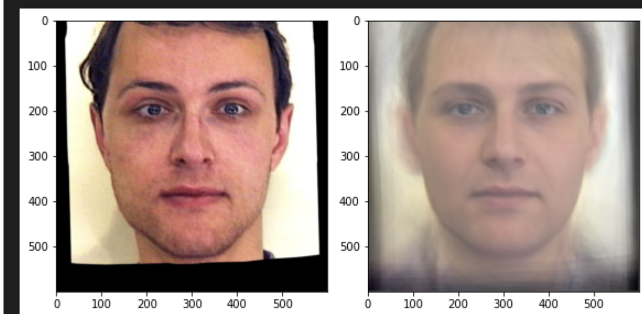


(1%) 請從數據集中挑出任意五張圖片，並用上題前五大 Eigenfaces 進行 reconstruction，並畫出結果。

Picture number 182

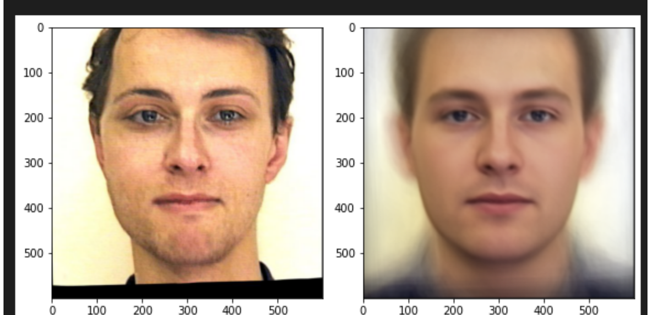


Picture number 288

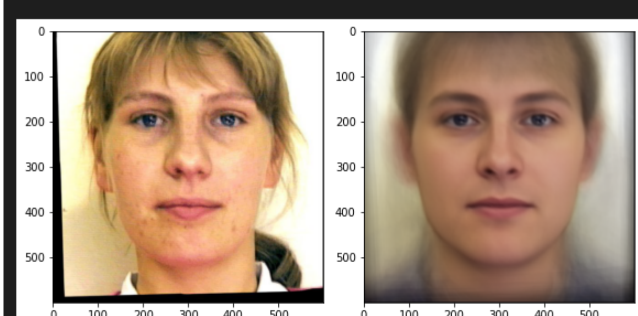


Picture number 140

Picture number 282



Picture number 398



EEML HW5 MATH

Problem1

(a)

1、

params to be estimated: $\theta = (\pi_k, \tau_k)_{k=1}^{2}$

likelihood function for the exponential mixture model.

$$p(X;\theta) = \prod_{i=1}^{3} p(x_i;\theta) = \prod_{i=1}^{3}\sum_{k=1}^{2}\pi_k f_\tau^k(x_i) = \prod_{i=1}^{3}\sum_{k=1}^{2}\pi_k \frac{1}{\tau_k} e^{-\frac{x_i}{\tau_k}}, \quad x_i \ge 0$$

$$\log p(X;\theta) = \sum_{i=1}^{3} \underbrace{\log\left(\sum_{k=1}^{2}\pi_k \frac{1}{\tau_k} e^{-\frac{x_i}{\tau_k}}\right)}_{\log P(x_i;\theta)}$$

(a)

In E-step, we need to derive

$$Q(\theta|\theta^{(t)}) = \sum_{z} p(z|X;\theta^{(t)}) \log p(X,z;\theta)$$

$$= \sum_{i=1}^{3} \mathbb{E}_{z\sim p(\cdot|X;\theta^{(t)})} \log p(x_i, z_i;\theta)$$

$$= \sum_{i=1}^{3} \mathbb{E}_{z_i\sim p(\cdot|x_i;\theta^{(t)})} \left[\log p(x_i, z_i;\theta)\right]$$

compute $p(z_i=j|x_i;\theta^{(t)}) = \dfrac{P(z_i=j, x_i;\theta^{(t)})}{\sum_{k=1}^{2} p(z_i=k, x_i;\theta^{(t)})}$

$$= \frac{\pi_j^{(t)} f(x_i;\tau_j^{(t)})}{\sum_{k=1}^{2}\pi_k^{(t)} f(x_i;\tau_k^{(t)})} \equiv h_{ij}^{(t)}$$

$$\Rightarrow Q(\theta|\theta^{(t)}) = \sum_{i=1}^{3}\sum_{j=1}^{2} P(z_i=j|x_i;\theta^{(t)}) \log P(x_i, z_i=j;\theta)$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)} \log \pi_j \frac{1}{\tau_j^{(t)}} e^{-\frac{x_i}{\tau_j}}.$$

$\frac{1}{\tau_j} = \lambda_j$

$$= \sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)} \log \pi_j \lambda_j e^{-\lambda_j x_i}.$$

(b)

(b)

In M-step, we compute $\theta^{(t+1)} = \underset{\theta}{\mathrm{argmax}}\ Q(\theta|\theta^{(t)})$

$$\Rightarrow \theta^{(t+1)} = \underset{\theta}{\mathrm{argmax}} \sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)} \log(\pi_j \lambda_j e^{-\lambda_j x_i})$$

$$= \underset{\theta}{\mathrm{argmin}} \sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)} (\lambda_j x_i - \log(\pi_j \lambda_j))$$

$$\text{s.t. } \sum_{j=1}^{2} \pi_j = 1$$

The lagragian $\mathcal{L} = \sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)} (\lambda_j x_i - \log\pi_j - \log\lambda_j) + \ell(\sum_{j=1}^{2}\pi_j - 1)$

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = \sum_{i=1}^{3} h_{ij}^{(t)} (x_i - \frac{1}{\lambda_j}) = 0 \quad -①$$

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = -\sum_{i=1}^{3} \frac{h_{ij}^{(t)}}{\pi_j} + \ell = 0 \quad -②$$

$$\frac{\partial \mathcal{L}}{\partial \ell} = \sum_{j=1}^{2} \pi_j - 1 = 0 \quad -③$$

$$② \Rightarrow \pi_j \ell = \sum_{i=1}^{3} h_{ij}^{(t)}$$

$$\Rightarrow \sum_{j=1}^{2} \pi_j \ell = \sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)}$$

By ③ $\Rightarrow \ell = \sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)}$ #. plug into

$$\Rightarrow \pi_j = \frac{\sum_{i=1}^{3} h_{ij}^{(t)}}{\sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^{(t)}} \quad \#$$

$$① \Rightarrow \lambda_j = \frac{\sum_{i=1}^{3} h_{ij}^{(t)}}{\sum_{i=1}^{3} h_{ij}^{(t)} x_i} \quad \left(\tau_j = \frac{1}{\lambda_j} = \frac{\sum_{i=1}^{3} h_{ij}^{(t)} x_i}{\sum_{i=1}^{3} h_{ij}^{(t)}} \right) \quad \#.$$

(c)

計算過程: 包含在MATH.ipynb內

i	**1**	**2**	**3**
x_i	0	0.2	4
$p(x_i, z_i = 1; \theta^{(t)})$	0.5	0.06767	0.00916
$p(x_i, z_i = 2; \theta^{(t)})$	0.25	0.09197	0.03383
$\sum_{j=1}^{2} p(x_i, z_i = j; \theta^{(t)})$	0.75	0.15964	0.04299
$\delta_{i1}^{[t]} = \mathbb{P}[z_i = 1 \| x_i; \theta^{[t]}]$	0.66667	0.42388	0.21301
$\delta_{i2}^{[t]} = \mathbb{P}[z_i = 2 \| x_i; \theta^{[t]}]$	0.33333	0.57612	0.78697

(d)

計算過程: 包含在MATH.ipynb內

$\tau_1 = 1.30398$, $\tau_2 = 2.53483$

$\pi_1 = 0.43452$, $\pi_2 = 0.56548$

Problem2

計算過程: 包含在MATH.ipynb內

(a)

```
Principle axes:
1-th principle axes: [-0.6165947  -0.58881629 -0.52259579]
2-th principle axes: [ 0.67817891 -0.73439013  0.02728563]
3-th principle axes: [ 0.39985541  0.33758926 -0.85214385]
```

(b)

```
Principle components:
Sample 1, principle components: [ 7.18658682  1.37323947 -2.25104047]
Sample 2, principle components: [ 0.75871342 -0.94399334 -0.73022635]
Sample 3, principle components: [-3.07034019 -4.45059025 -3.1883001 ]
Sample 4, principle components: [ 2.60849751 -2.97853006 -1.92979259]
Sample 5, principle components: [-1.82299166 -4.75401212  4.25159619]
Sample 6, principle components: [3.35457763 3.91896138 2.52755823]
Sample 7, principle components: [-4.41464321  2.55604371 -2.13952468]
Sample 8, principle components: [ 3.46569126 -1.73131477  2.27849363]
Sample 9, principle components: [-2.31359638  6.03371503  0.2038499 ]
Sample 10, principle components: [-5.75249521  0.97648096  0.97738622]
```

(c)

```
Reconstruction dataset
[[ 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]]
```

Average reconstruction error: 5.47203

Problem3

(a)

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

show symmetric:

① $(AA^T)^T = (A^T)^T A^T = AA^T$

② $(A^TA)^T = (A^T)(A^T)^T = A^TA$

show positive semi-definite:

$\forall x \in \mathbb{R}^n,\ x \neq \vec{0}$

① $x^T(AA^T)x = (A^Tx)^T(A^Tx) = \|A^Tx\|^2 > 0$

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

show they share the same eigenvalue:

Let x be some non-zero eigenvector of AA^T with eigenvalue $\lambda \neq 0$

$\Rightarrow (AA^T)x = \lambda x$.

左乘 A^T

$\Rightarrow (A^TA)\underbrace{(A^Tx)}_{y} = \lambda \underbrace{(A^Tx)}_{y}$, $y \neq \vec{0}$

$\Rightarrow (A^TA)y = \lambda y$, which means that A^TA has the same eigenvalue λ #.

(b)

(b)

$\Sigma \in \mathbb{R}^{m \times m}$, sym and p.s.d.

$u \in \mathbb{R}^m$.

$\because \Sigma$ is p.s.d and symmetric

$\therefore \Sigma = LDU$ (LDU decomposition)

$= LDL^T$ (Σ is sym.)

$= \sum_{i=1}^{n} (d_i l_i l_i^T)$

Assume $u = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^m$ for simplicity.

Let $n = 2m$, and construct a set of points $x_1, \cdots, x_m, x_{m+1}, \cdots x_{2m}$

where $x_i = \sqrt{d_i}\, l_i \in \mathbb{R}^m, \forall i,\ 1 \leq i \leq m$

$x_i = -\sqrt{d_i}\, l_i \in \mathbb{R}^m, \forall i,\ m+1 \leq i \leq 2m$.

$\frac{1}{n}\sum_{i=1}^{n} x_i = u = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

$\frac{1}{n}\sum_{i=1}^{n} (x_i - u)(x_i - u)^T = \sum_{i=1}^{n} d_i l_i l_i^T = LDL^T = \Sigma$.

(c)

(c)

we know that $\frac{\partial}{\partial X} Tr(X^T A X) = (A+A^T)X$.

$$\min_{\Phi} Tr(\Phi^T \Sigma \Phi) \quad s.t. \quad \Phi^T \Phi = I$$

$$\mathcal{L} = Tr(\Phi^T \Sigma \Phi) - \lambda(\Phi^T \Phi - I)$$

$$\frac{\partial \mathcal{L}}{\partial \Phi} = (\Phi + \Phi^T)\Sigma - 2\lambda \Phi = 0$$

$\Phi = \Phi^T$

$\Rightarrow \Sigma \Phi = \lambda \Phi$.

$\Rightarrow \Phi = (\Sigma - \lambda)^{-1}$