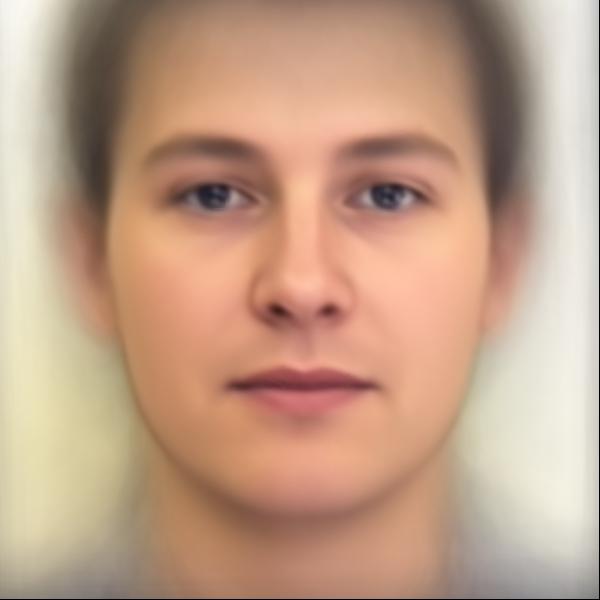
Machine Learning HW7 Report

學號：R07943095 系級：電子所碩一 姓名：劉世棠

1. PCA of color faces:
2. 請畫出所有臉的平均。



1. 請畫出前五個 Eigenfaces，也就是對應到前五大 Eigenvalues 的 Eigenvectors。



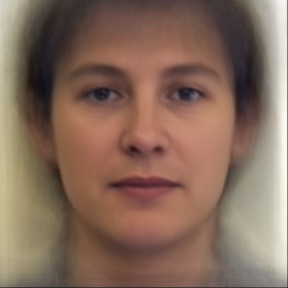
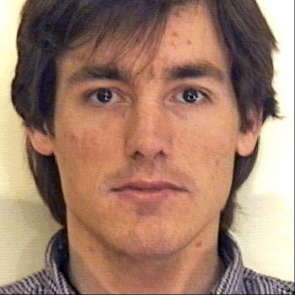
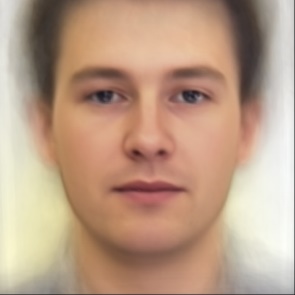
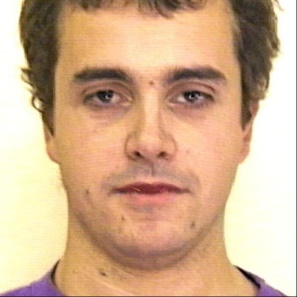
1 2 3

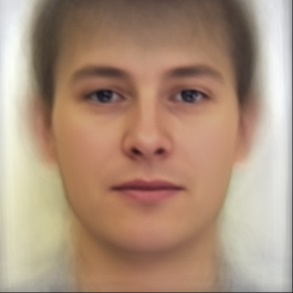
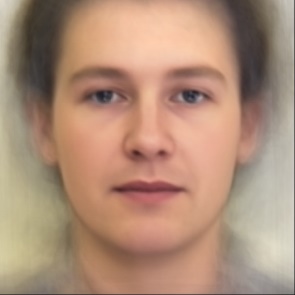


4 5

**這題我重建可能會與助教不同，因為eigenvector的可以有正反，所以會造成與助教重建有的有些不同(shuffle過後用pca的方向可能會跟原本相反)。**

1. 請從數據集中挑出任意五張圖片，並用前五大 Eigenfaces 進行 reconstruction，並畫出結果。

 1.jpg  10.jpg

 22.jpg  37.jpg

 72.jpg

**這題用助教的環境跑過跟同學測試相同。**

1. 請寫出前五大 Eigenfaces 各自所佔的比重，請用百分比表示並四捨五入

到小數點後一位。

**依照排序前五大Eigenfaces分別為: 4.1%、2.9%、2.4%、2.2%、2.1%**

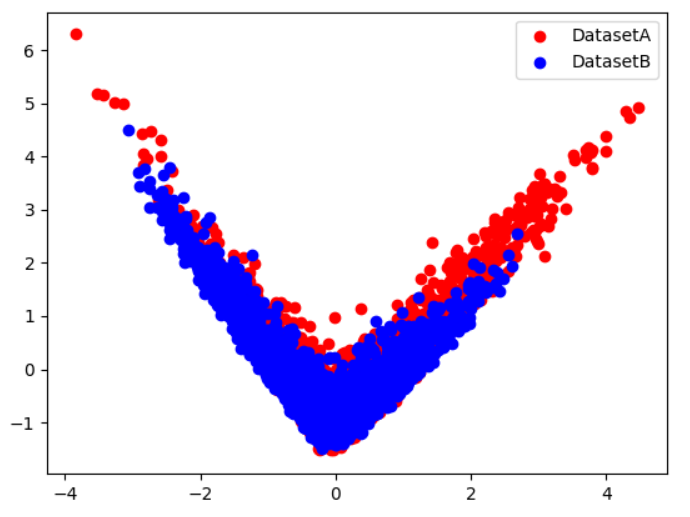
2. Image clustering:

1. 請實作兩種不同的方法，並比較其結果(reconstruction loss, accuracy)。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

這題我是實作autoencoder + pca和autoencoder only兩種做法，故我的reconstruction loss(mae)都是一樣約為0.03224，而有經過PCA(有開啟whiten)的準確率大幅提升(直接衝破simple和strong)，以下為圖表:

|  |  |  |
| --- | --- | --- |
|  | public | private |
| Autoencoder only | 51.063% | 51.337% |
| Autoencoder+pca | 97.281% | 97.263% |

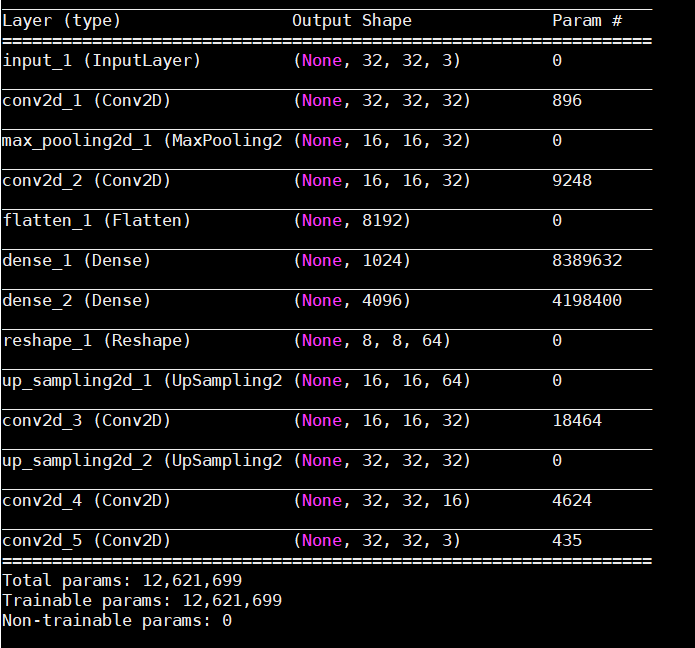
1. 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。  
   (用 PCA, t-SNE 等工具把你抽出來的 feature 投影到二維，或簡單的取前兩維2的 feature)   
   其中visualization.npy 中前 2500 個 images 來自 dataset A，後 2500 個 images 來自 dataset B，比較和自己預測的 label 之間有何不同。



**以上為我用autoencoder+PCA降成64維取前兩維的結果，可能是投影的時候並未取到重要的維度，所以看起來光靠前兩維並不能好好地分開資料，而後來我將64維的資料丟入Kmeans，最後得到98.64%的精準度。**

c. 請介紹你的model架構(encoder, decoder, loss function...)，並選出任意

32張圖片，比較原圖片以及用decoder reconstruct的結果。



**介紹:**

**以上為我所使用的autoencoder架構，loss funcion是mae，除此之外我還會使用PCA降維至96維，經過這次作業發現PCA實在強大，經過PCA的降維(whiten=true)後做Kmeans，kaggle上的成績直接從51%變成96%。**

**以下為32~63的圖片(上排是原圖，下排是reconstruction)**

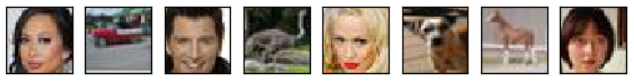
**Original:**



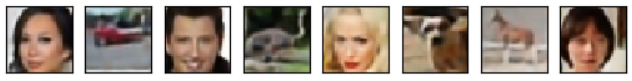
**Reconstruction:**



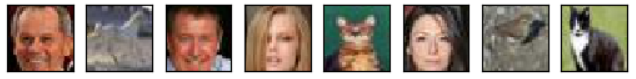
**Original:**



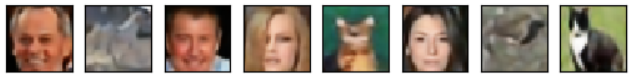
**Reconstruction:**



**Original:**



**Reconstruction:**



**Original:**



**Reconstruction:**

