

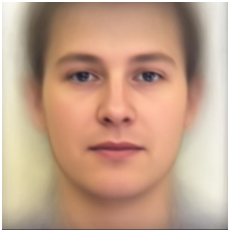
Machine Learning HW7 Report

學號：B06705057 系級：資工二

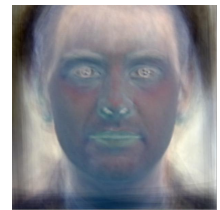
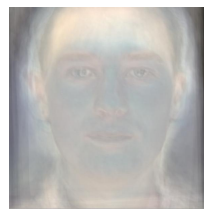
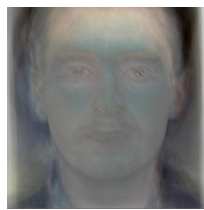
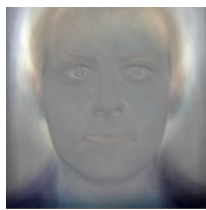
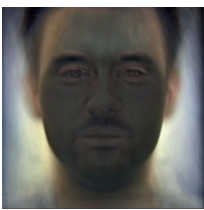
姓名：黃資翔

1. PCA of color faces:

a. 請畫出所有臉的平均。

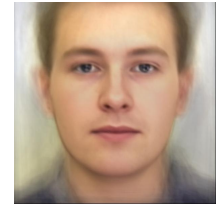
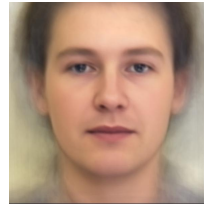
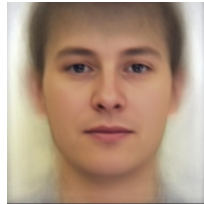
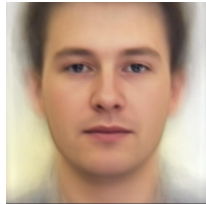
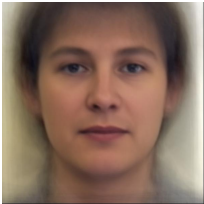


b. 請畫出前五個 Eigenfaces，也就是對應到前五大 Eigenvalues 的 Eigenvectors。
Eigenvalues 越左邊越大



c. 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。

從左至右分別為圖 1, 10, 22, 37, 72



d. 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。

4.1%	2.9%	2.4%	2.2%	2.1%
------	------	------	------	------

2. Image clustering:

a. 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)

第一種方法 降維方法：純 PCA

cluster 方法：2-mean

第二種方法 降維方法：auto-encoder -> PCA

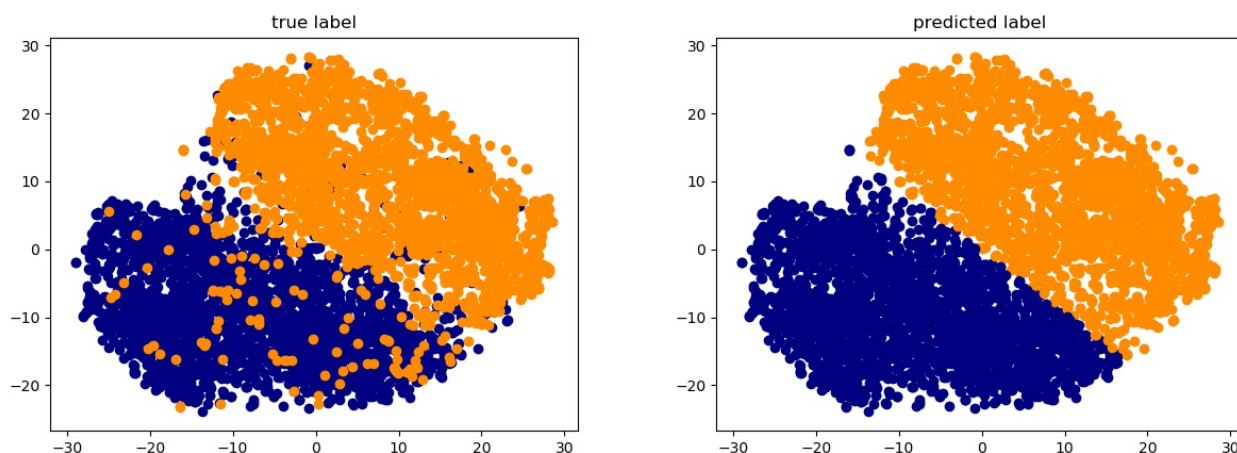
cluster 方法：2-mean

Reconstruction loss : root mean square error

	Reconstruction loss	Public accuracy	Private accuracy
方法一	2.469	0.95518	0.95481
方法二	3.924	0.96623	0.96590

b. 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。

利用 t-SNE 將圖片投影至二維，藍色為 dataset A，橘色為 dataset B



由圖片可以看出 K-means 在二維時效果蠻好，並且能精確的找出兩類別的分界。然而預測錯誤的關鍵是降維不夠好，或者是僅用二個維度無法完全區分兩類圖片。

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。

encoder : Conv2D nuerons = 64 -> Conv2D nuerons = 8 -> PCA components = 300
 Decoder : PCA inverse -> Conv2D nuerons = 8 -> Conv2D nuerons = 64 -> Conv2D nuerons = 3
 Loss function : MSE

