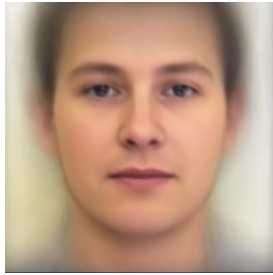


HW4

學號：R06725041 系級：資管碩一 姓名：彭証鴻

A. PCA of colored faces

A.1. (.5%) 請畫出所有臉的平均。

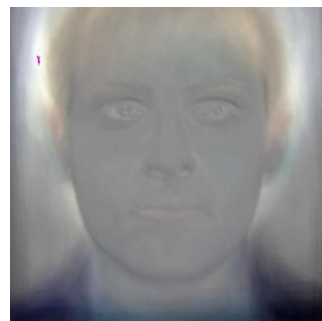


A.2. (.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。

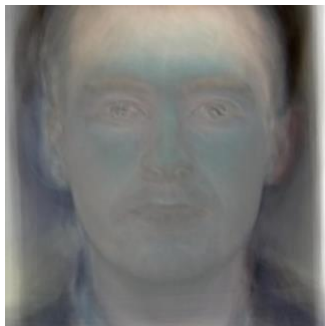
Eigenface1



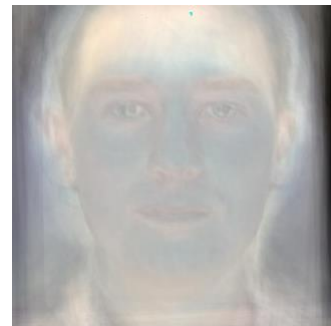
Eigenface2



Eigenface3



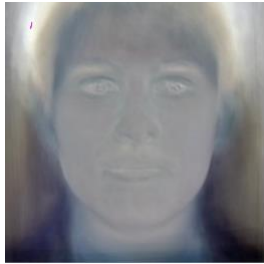
Eigenface4



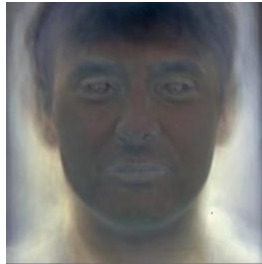
A.3. (.5%) 請從數據集中挑出任意四個圖片，並用前四大 Eigenfaces 進行 reconstruction，並畫出結果。

下圖為從數據集中挑選 index 為 110, 200, 300 和 400 的四張圖片來 reconstruct，結果如下：

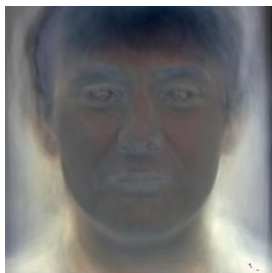
110.jpg



200.jpg



300.jpg



400.jpg



A.4. (.5%) 請寫出前四大 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。

1. 4.1%

2. 2.9%

3. 2.4%

4. 2.2%

B. Image clustering

B.1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

Model1 => AutoEncoder + Kmeans

public score : 0.99971 private score : 0.99974

Model2 => PCA + Kmeans

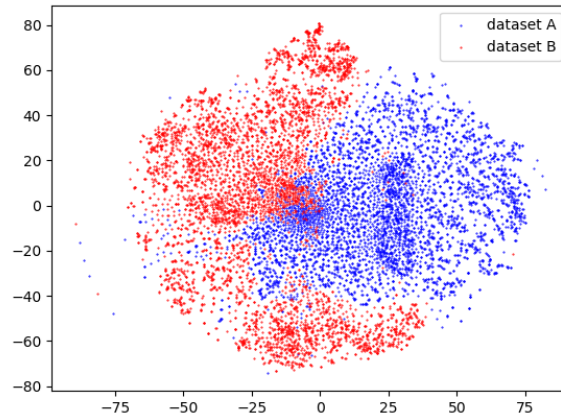
public score : 1 private score : 1

B.2. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。

透過由 B.1.的 PCA Model 先降至 300 維，隨後再用 tsne 投影至二

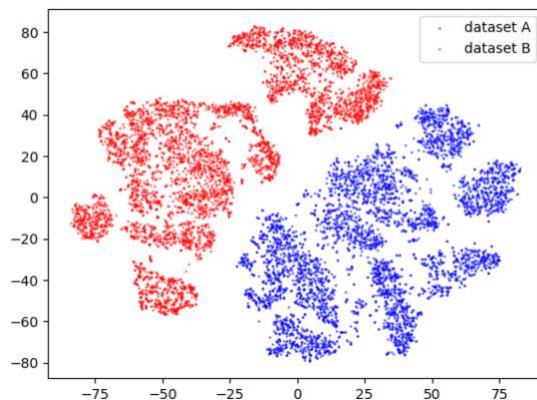
維做圖的結果如下

Predict label:



B.3. (.5%) visualization.npy 中前 5000 個 images 跟後 5000 個 images 來自不同 dataset。請根據這個資訊，在二維平面上視覺化 label 的分佈，接著比較和自己預測的 label 之間有何不同。

Answer:



觀察上面 answer 與第二小自己預測的 label 後，發現儘管我用第一小題 public score =1.0 的 PCA model 下去降維，一樣沒有辦法分的很乾淨，上網查了一下可能是因為 PCA 的原理有關。因為 PCA 是對資料求共變異數矩陣，在進行奇異值分解。因此會被資料的差異性影響，無法很好表現相似性以及分佈。

C. Ensemble learning

C.1. (1.5%) 請在 hw1/hw2/hw3 的 task 上擇一實作 ensemble learning ,

請比較其與未使用 ensemble method 的模型在 public/private score 的表現並詳細說明你實作的方法。(所有跟 ensemble learning 有關的方法都可以，不需要像 hw3 的要求硬塞到同一個 model 中) (Collaborators: 參考網路上資料，網址如下：
<https://goo.gl/c9L2CR>)

這一題我選擇 hw3 的 task 來實作 ensemble learning，我透過 stacking 的方法將我四個不同 model 的 output 做 Stacking(Voting)，其中 3 個 model 完全照著網路上的 tutorial 建立，而另一 model 為我在 hw3 的 simple model。

基本參數設定：num_epochs = 20, batch_size = 32。

以下是單一 model 和 ensemble model 在 val_error 上表現的比較：

單一 model val_error: 1.0865 local val_acc = 0.6071

ensemble mdoel val_error: 0.9999 local val_acc = 0.6360

依照實驗所呈現結果，確實觀察到 ensemble model 無論是在 val_loss 有下降的現象，而 val_acc 則有所提升。