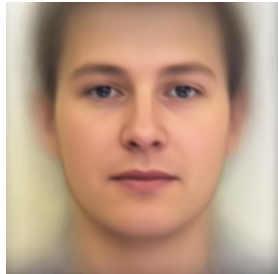


HW4

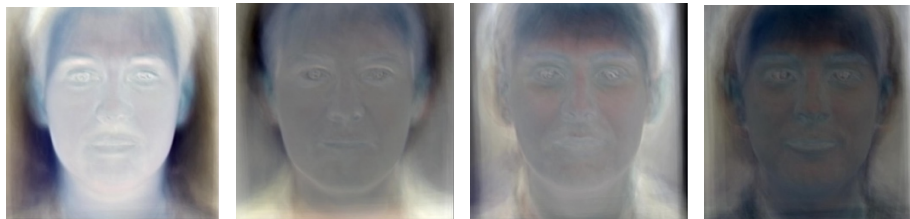
學號：B04502139 系級：電機三 姓名：戴瑋辰

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。



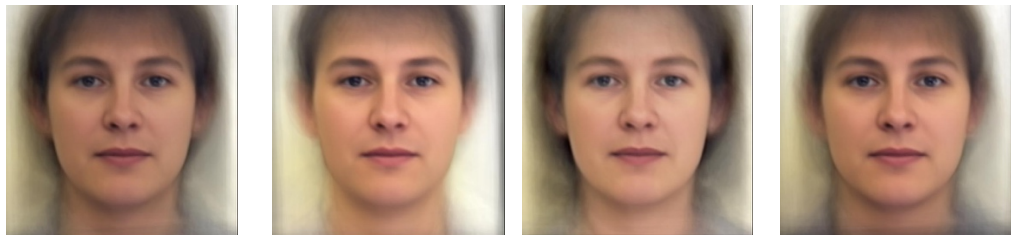
Eigenface 1

Eigenface 2

Eigenface 3

Eigenface 4

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。



Reconstruct 1.jpg

Reconstruct 100.jpg

Reconstruct 250.jpg

Reconstruct 300.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。

Eigenface	1	2	3	4
Proportion	4.1%	2.9%	2.4%	2.2%

B. Image clustering

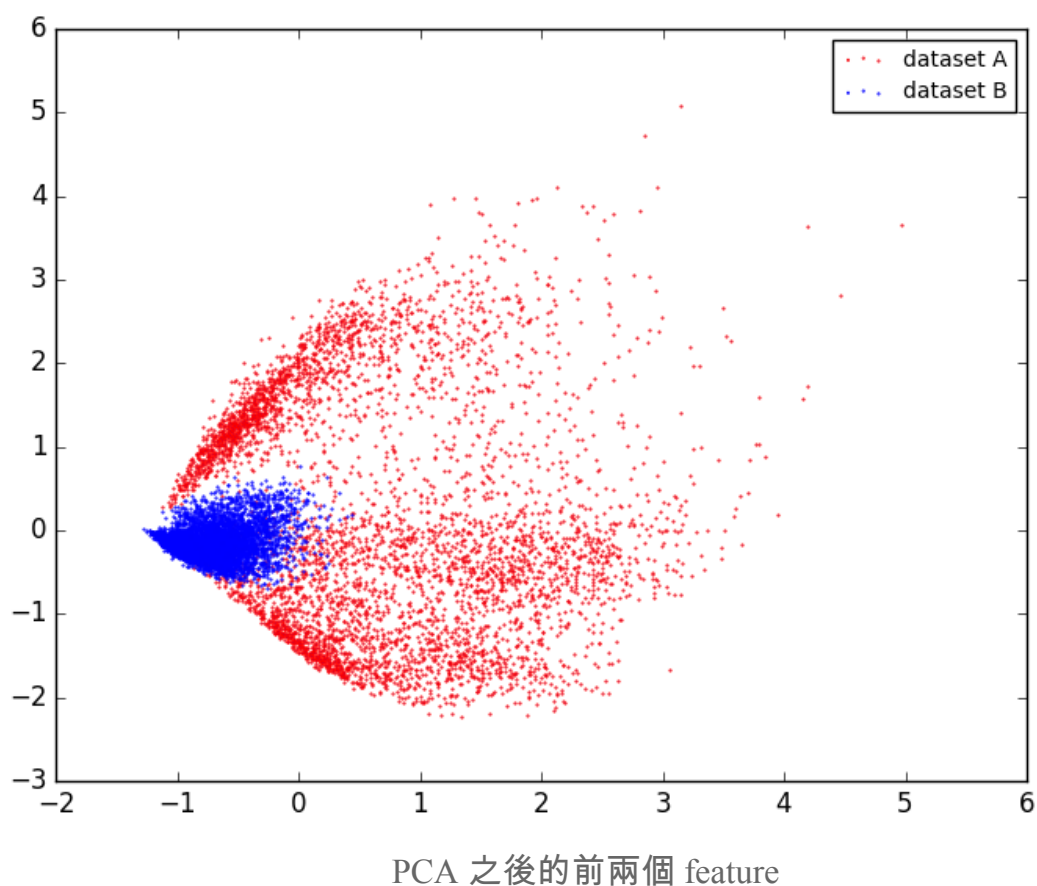
- B.1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

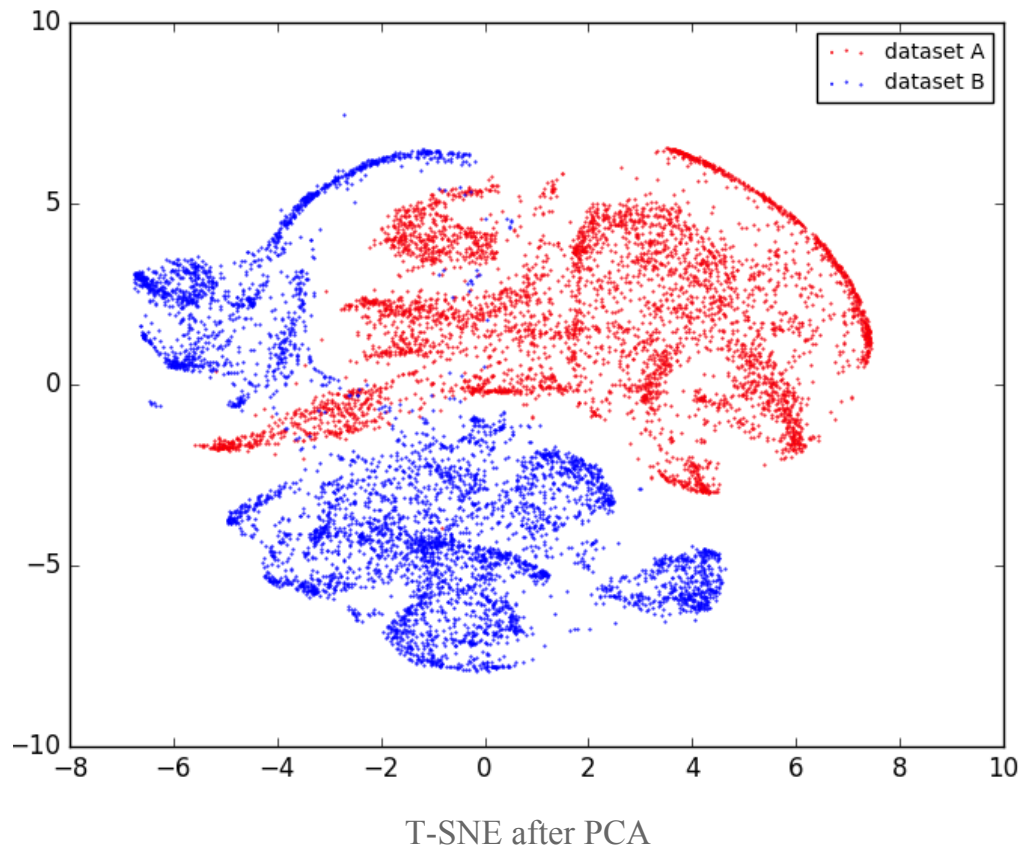
Dimension reduction	<i>PCA</i>	<i>Auto encoder</i>
Clustering	<i>KMeans</i>	<i>KMenas</i>
Score on Kaggle	100%	100%

實做 PCA 時，我設定保留 99% 的 variance，最後將 784 維的資料縮到 567 維，最後用 kmeans 預測資料來源，準確率 100%

而 Auto encoder 則是用 convolutional Auto encoer 實作，encode、decode 對稱，各三層；將原本 784 為的資料縮到(4, 4, 8)，也就是 128 維，降維完成後一樣用 kmeans 預測，得到 100% 的正確率。

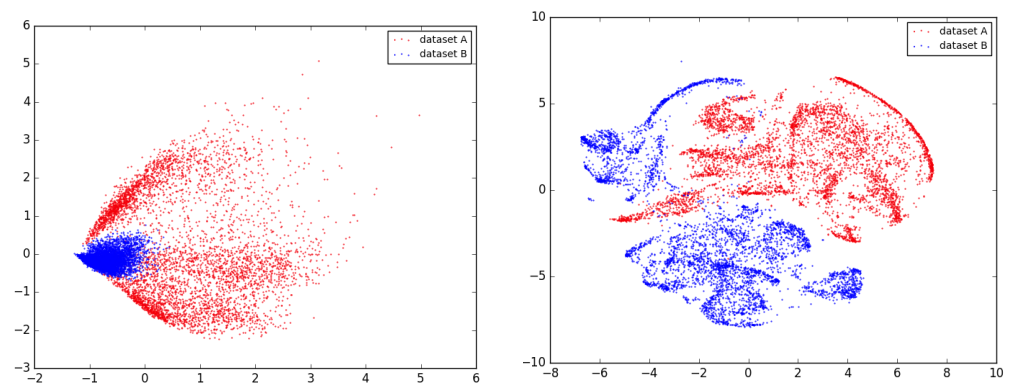
- B.2. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。





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。

由我的模型預測出的 label 和結果一樣，因此圖片完全相同。



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。

我選擇使用 hw3 的 model 來做 ensemble learning。在實作之前，我先
以不同的參數 train 了 4 個 model，其在我自己所切的 validation set 上準
確率分別為 66.3%, 65.9%, 65.8%, 67.2%。

接著我把這四個 model 合在一起，作法為將 softmax 的輸出先全部相加
取平均，再從這幾個平均的機率中選出一個最大的。

Kaggle Score

	Model 1	Model 2	Model 3	Model 4	Ensemble
Public	0.6623	0.6498	0.6539	0.6723	0.7027
Private	0.6765	0.6517	0.6673	0.6715	0.6918

從表中可以看到在做了 ensemble 之後，表現明顯提升很多。