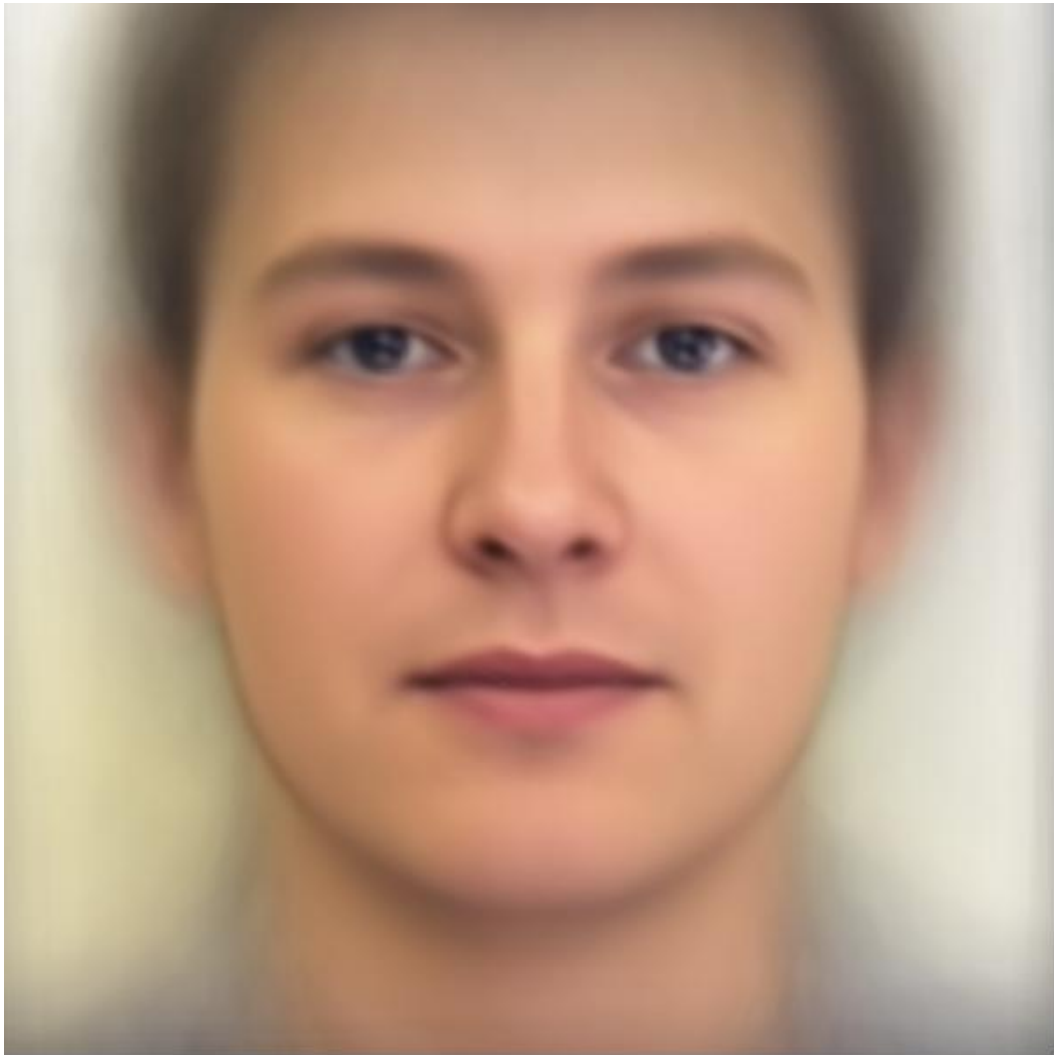


ML2018SPRING – HW4 REPORT

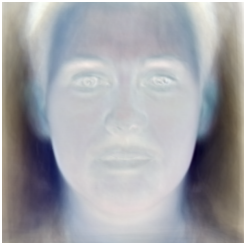

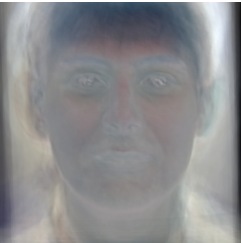
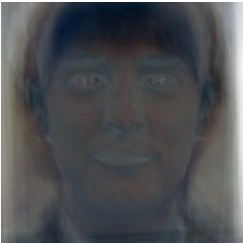
學號：B05902052 系級：資工二 姓名：劉家維

1. PCA of colored faces

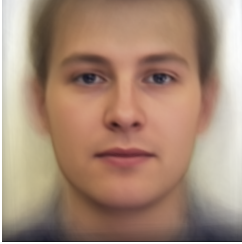
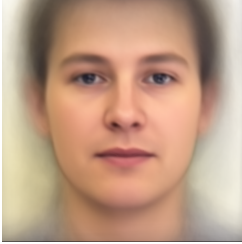
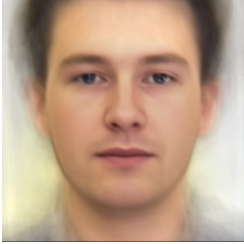
1. (.5%) 請畫出所有臉的平均。



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

rank	1	2	3	4
picture				

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。

index	0	3	6	9
picture				

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。

Ans: 4.1%, 2.9%, 2.4%, 2.2%

2. Image clustering

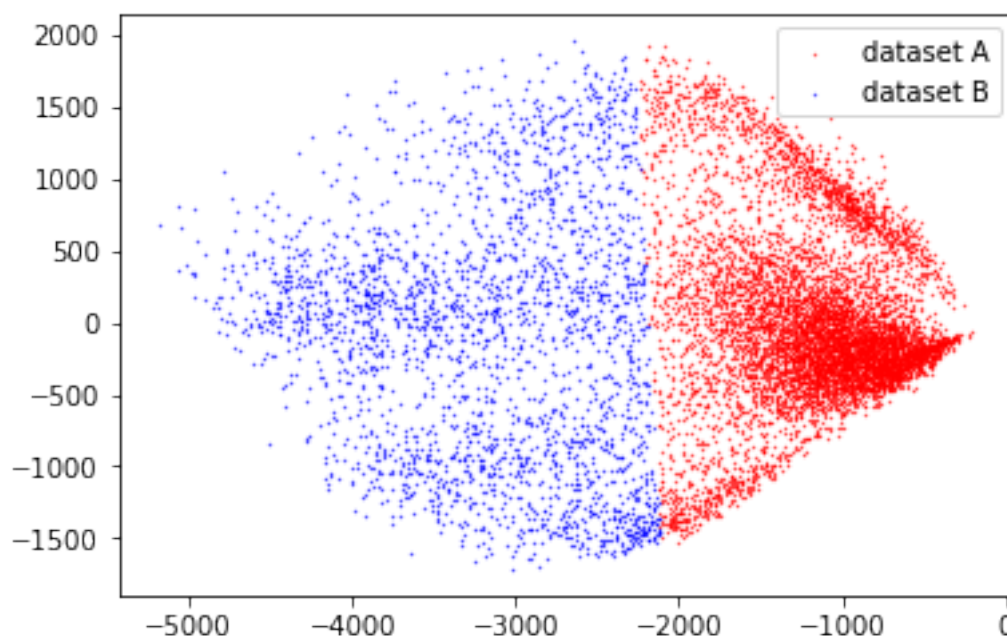
1. (.5%) 請比較至少兩種不同的 feature extraction 及其結果。
(不同的降維方法或不同的 cluster 方法都可以算是不同的方法)

SVD 降維 + k-means clustering:

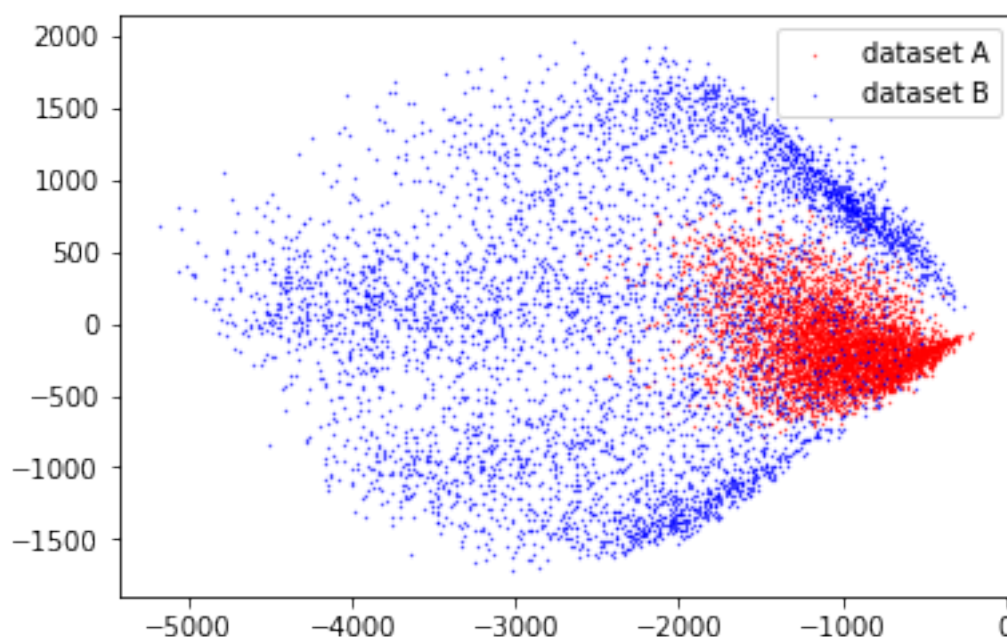
T-SNE 降維 + k-means clustering:

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

實作方法：將資料 SVD 降維後取前兩維 feature，用 k-means 得出的兩個中心為準分出兩個 cluster。



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。



透過圖表可以發現，用 SVD 降維後取前兩維的方式其實並不能將資料集分得很開，

雖然紅點在降維後很集中，但是藍點散布在整個圖中，而我做 clustering 的方式又是 k-means，如此造成兩個 clustering 大小有著非常大的差異（2.圖表中藍點為 7444 個，紅點則為 2556 個，但其實兩個 cluster 應該要各 5000 個）

3.Ensemble learning

- 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）

Task: hw2

Models:

- (1) logistic regression
- (2) support vector classifier
- (3) neural network

Layer (type)	Output Shape	Param #
=====	=====	=====
dense_1 (Dense)	(None, 2048)	251904
dropout_1 (Dropout)	(None, 2048)	0
dense_2 (Dense)	(None, 256)	524544
dropout_2 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 1)	257

用以上三個模型進行預測，看哪項預測結果佔多數（2 個模型以上預測為 0 即為 0，2 個模型以上預測為 1 即為 1），就選擇該預測結果。

準確率：(public / private)

Logistic Regression: 0.85503 / 0.84707

Support Vector Classifier: 0.85724 / 0.84645

Neural Network: 0.84398 / 0.84719

Ensemble:

0.86031 / 0.85186