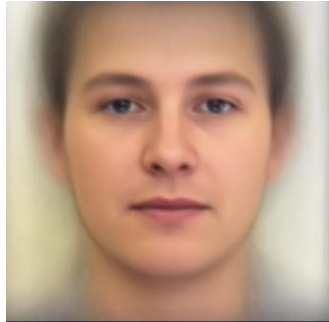


## HW4




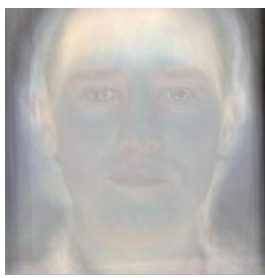
學號：r05323040 系級：經濟碩二 姓名：田家駿

### A. 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。

Eigenvalues 0	Eigenvalues 1	Eigenvalues 2	Eigenvalues 3
			

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。

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%

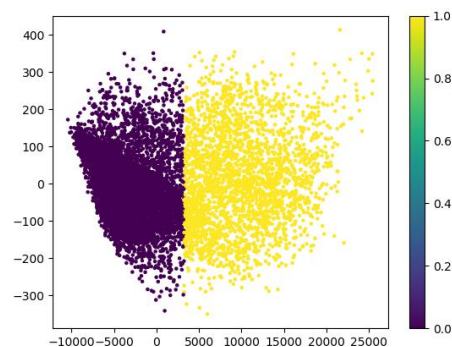
## B. Image clustering

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)

兩種方法分別為 PCA 降維到 30 feature 再去做 kmeans 分兩種以及使用 autoencoder 降維到 30 feature 再去做 kmeans，autoencoder 的架構參考助教手把受教學，這裡 epoch 為 100 次，上傳 kaggle 的成績 Pca 為 0.51696，autoencoder 為 0.51641，可以看到用更複雜的方式去降維在這筆資料中並沒有比較好，也有可能是 autoencoder 還未收斂到 global minimum。

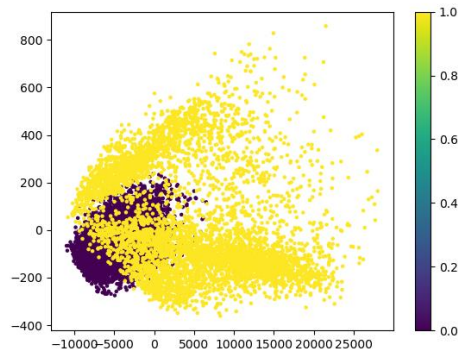
2. (.5%) 預測 visualization.npy 中的 label，在二維平面上視覺化 label 的分佈。

這裡是用上題的 autoencoder 及 kmeans 來預測，因 tsne 時間很久，這裡是再使用 PCA 降到二維平面上。



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。

我們可以看到 kmeans 比較有問題的地方在兩個顏色重疊的地方，我的預測把左邊黃色的地方都認定為藍色。



## C. 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。這裡的 ensemble model 我使用的是 random forest，再 random forest 中，我們需要去 tune 樹的深度，這裡我 validation 出來深度是 20，並用 500 個 estimator 去決定，上傳的分數為下表，給定一樣深度為 20 的 decision tree，ensemble 的方法增加了 3 percent 的預測率。

[decision.csv](#)

a few seconds ago by [r05323040\\_台大經濟田家駿](#)

[add submission details](#)

0.83552

0.83108



[ensem.csv](#)

7 minutes ago by [r05323040\\_台大經濟田家駿](#)

[add submission details](#)

0.86340

0.86547

