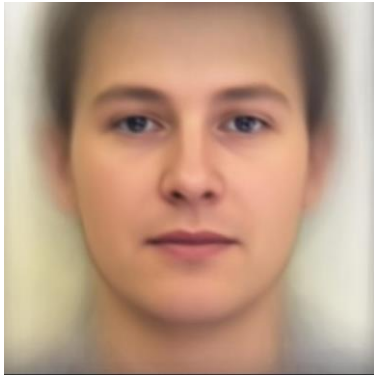


HW4

學號：r06521705 系級：土木系營管組碩一 姓名：陳思愷

A. PCA of colored faces

A.1. (.5%) 請畫出所有臉的平均。

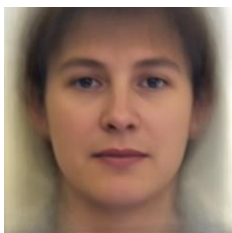


A.2. (.5%) 請畫出前四個 Eigenfaces，也就是對應到前四大 Eigenvalues 的 Eigenvectors。

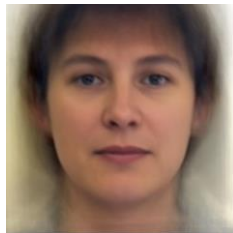


A.3. (.5%) 請從數據集中挑出任意四個圖片，並用前四大 Eigenfaces 進行 reconstruction，並畫出結果。

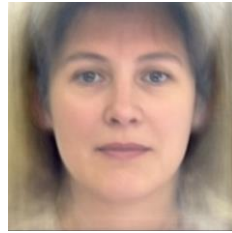
Num1:



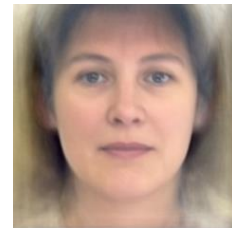
Num2:



Num3:



Num4:



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。

→4.1% 2.9% 2.4% 2.2%

B. Image clustering

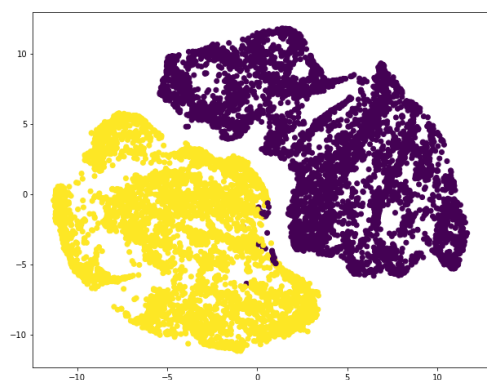
B.1. (.5%) 請比較至少兩種不同的 **feature extraction** 及其結果。(不同的降維方法或不同的 **cluster** 方法都可以算是不同的方法)

第一次我嘗試的是使用 **PCA** 降維到 32 維(**PCA(n_components=32)**)

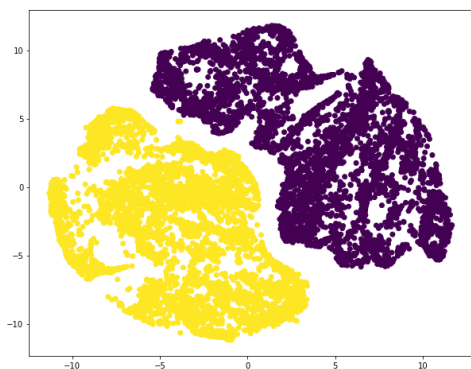
再將降維後的 **data** 丟到 **k-means** 裡面將其分成兩類。

第二次是嘗試使用 **auto-encoder** 的方式去將資料降維，一樣再丟到 **k-means** 裡面將其分成兩類。第一個方法在 **kaggle** 上的方法不太理想，結果只有 0.51716 分，但例外一個方法在 **kaggle** 可以拿到 0.97862 分，結果相差相當多。

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。



大致上的分類是正確的，作法是先利用 **autoencoder** 的方法將維度降到 32 維，再利用 **t-sne** 去進行投影到 2 維上的繪製。

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）

這次選用的是 hw2，收入是否大於 50k 那一份。

首先先用原本的 model(keras , logistic)跑出來的結果丟入 kaggle 紀錄分數:

	keras	Logistic
private	0.85652	0.84829
public	0.85724	0.85761

這次 ensemble 的做法是將上述兩個模型所得出來的結果

乘上各自被賦予權重 w , $X_{\text{keras}} * w_1 + X_{\text{logistic}} * w_2 + b = y$

去做 gradient decent 希望可以找出各自的權重以及 b

訓練結束後得出的權重:

$W = [5.56490382 , 0.813641]$ 顯示新 model 比較相信 keras 的結果

作後再把 test 資料丟入利用 ensemble 求出新的解丟入 kaggle

Ensemble:

Private: 0.85579

Public: 0.85982

不論是在 public 還是 private 都有些許的進步，進步幅度不夠可是因為利用到的 model 只有兩個，如果可以擴增多一點 model 一起進行可能會有更好的結果。