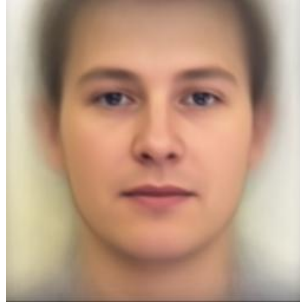


HW4

學號：r06521601 系級：土木碩一 姓名：黃伯凱

A. PCA of colored faces

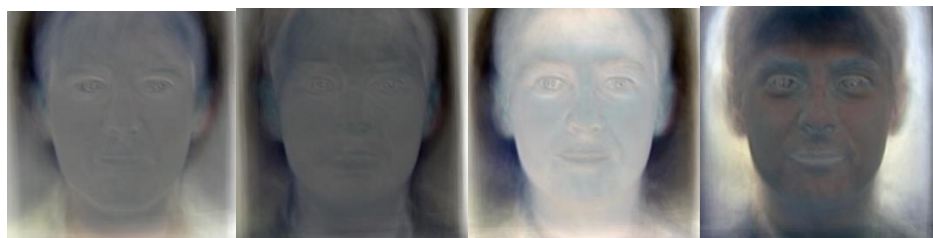
A.1. (.5%) 請畫出所有臉的平均。



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。



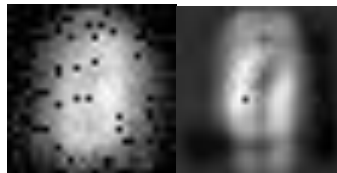
A.4. (.5%) 請寫出前四大 Eigenfaces 各自所佔的比重，請用百分比表示並四捨五入到小數點後一位。

針對前四大 eigenface 計算其 eigenvalue 除以總 eigenvalue 分別為：
4.1%、2.9%、2.4%、2.2%。

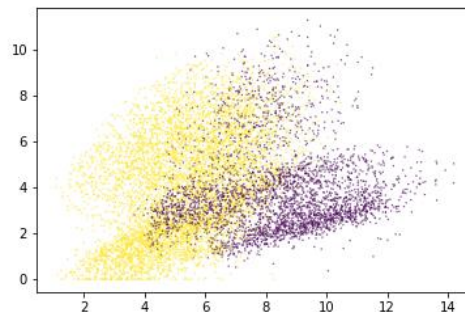
B. Image clustering

- 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)

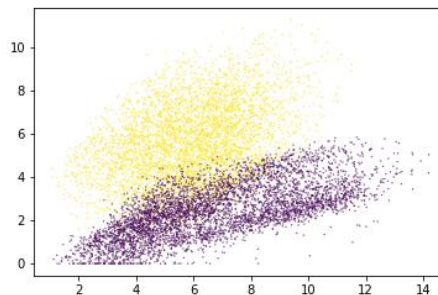
我這次使用了 autoencoder 以及 PCA 作為降維的方法，以下是我將兩種降維方法個別 reconstruction 回去後的結果(左為 autoencoder、右為 PCA)。兩種方法分別在 kaggle 上得到的分數是 autoencoder 較好的 98%及 PCA 得到 92%的準確率。



- 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。



這張圖是原始的 label 分佈，我發現和我第二題所預測的結果出現兩部份的誤差，分別是左下群以及右上群，我預測的方式為先使用 encoder 降維，再透過 kmeans 去分類。

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)

這題我選作業 2 作為我的實作案例, 我總共實做了三個模型, 分別為以下, 分別用不同的層數、activation 以及 optimizer 來製作模型:

```
model_1 = Sequential()
model_1.add(Dense(39, input_dim=39, activation='sigmoid'))
model_1.add(Dense(1, activation='sigmoid'))
model_1.compile(optimizer=Adam(lr=0.001),
                loss=binary_crossentropy,
                metrics=[binary_accuracy])

model_2 = Sequential()
model_2.add(Dense(39, input_dim=39, activation='sigmoid'))
model_2.add(Dense(14, activation='relu'))
model_2.add(Dense(1, activation='sigmoid'))
model_2.compile(optimizer=SGD(lr=0.01),
                loss=binary_crossentropy,
                metrics=[binary_accuracy])

model_3 = Sequential()
model_3.add(Dense(39, input_dim=39, activation='sigmoid'))
model_3.add(Dense(14, activation='relu'))
model_3.add(BatchNormalization())
model_3.add(Dropout(0.3))
model_3.add(Dense(1, activation='sigmoid'))
model_3.compile(optimizer=Adam(lr=0.001),
                loss=binary_crossentropy,
                metrics=[binary_accuracy])
```

三個模型分別驗證訓練模型的精度準確度為 1>3>2, 第二個模型的準確度較差, 所以我實作 ensemble 的方法為, 我將三者預測的結果整合成一個 list, 並且分配給三個模型不同的權重(因為 2 較差所以只有 0.8), 實作程式碼如下:

```
prediction_ensemble = np.average(prediction_list, axis=0, weights=[1,0.8,1])
```

最後上傳作業 2 的 kaggle 所得到分數比過去的 best 提高了 0.01, 看起來相當不錯的 young 子!