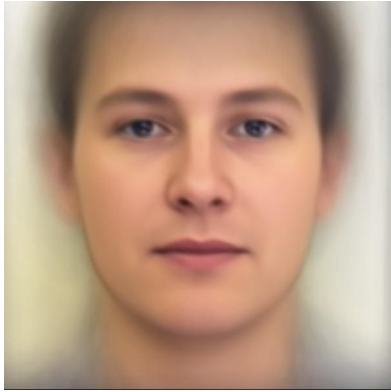
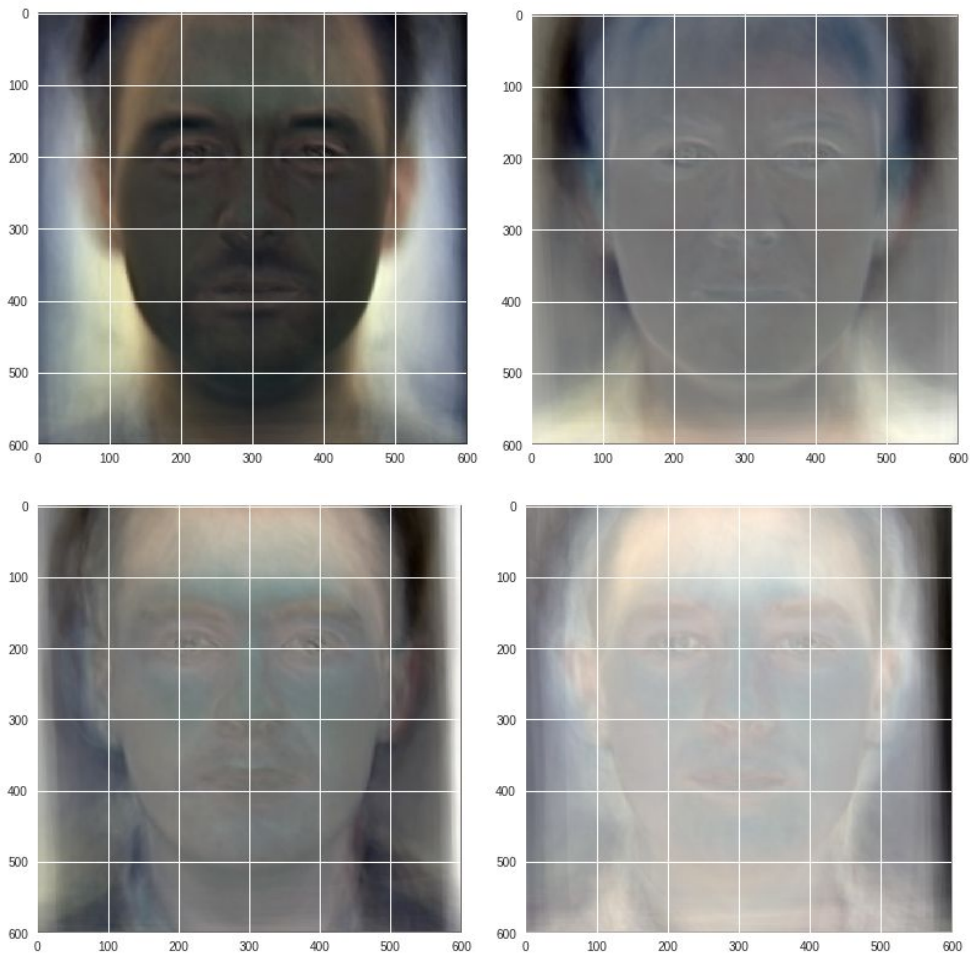


A. PCA of colored faces

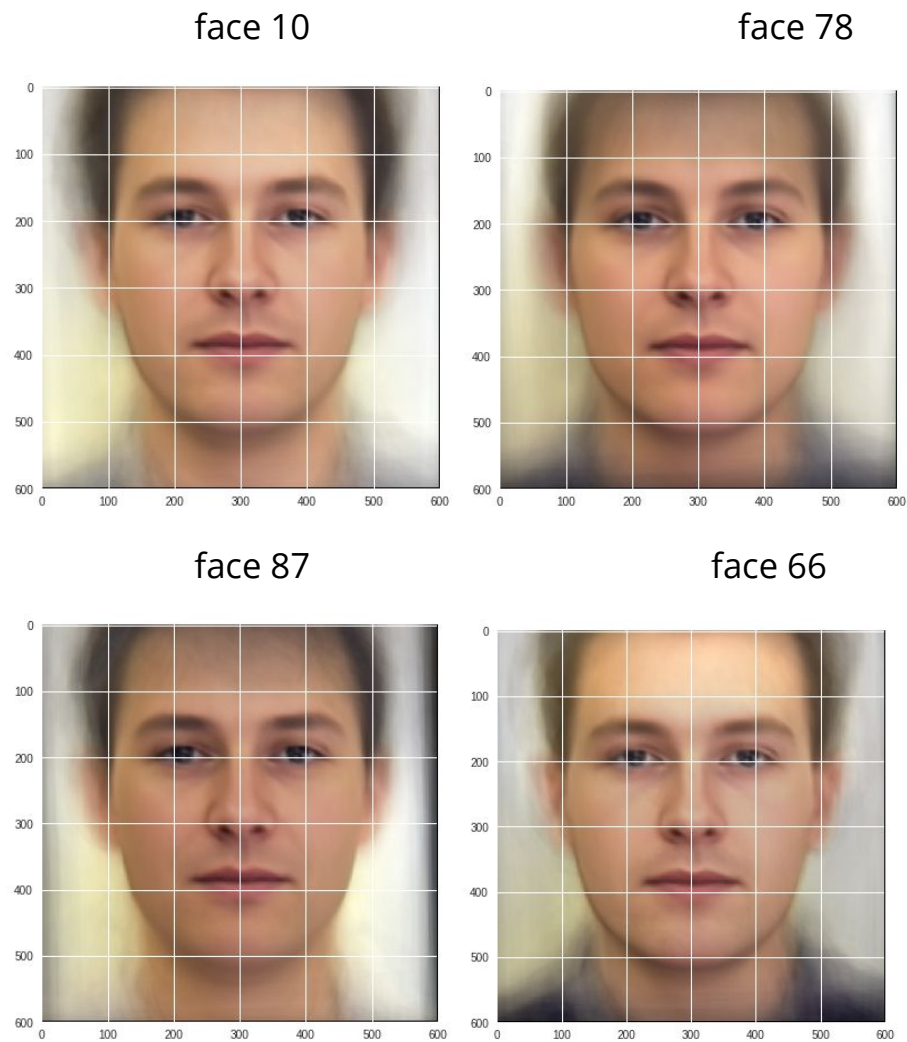
A.1. (.5%) 請畫出所有臉的平均。



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。

4.1%, 2.9%, 2.4%, 2.2%

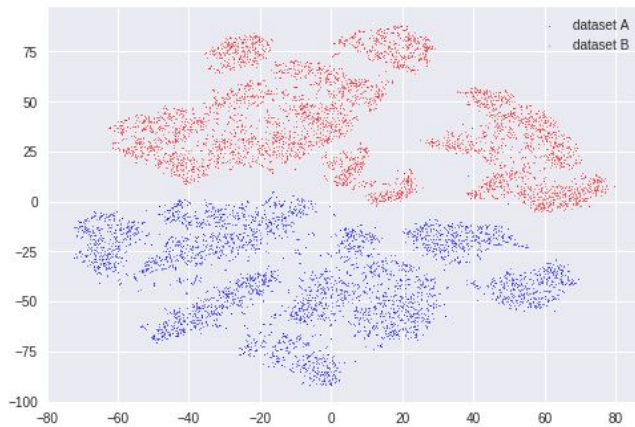
B. Image clustering

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)

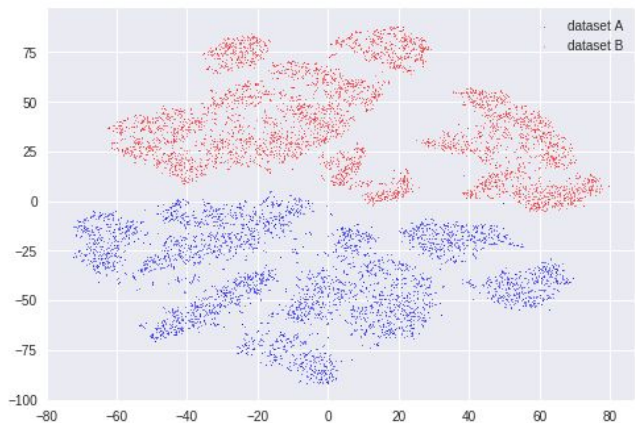
用sklearn的PCA+K-means的結果：public和private score都是0.99998

用PCA降維+t-sne+K-means結果：

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。



用TSNE降維投影的結果，跟直接取前5000個和後5000個前兩維投影的結果幾乎一模一樣，所以可見預測的結果應該算是蠻準的，可以把兩個label分得很開。

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）

取三個model做ensemble，取average
mode1: private: 0.67149 public: 0.67651
mode2: private: 0.67121 public: 0.68041
mode3: private: 0.66453 public: 0.67344
結果：private: 0.67762 public: 0.69211
比我kaggle原本最高分的還高，效果群拔