學號：r09922065 系級：資工所碩二 姓名：張庭逸

**Kaggle Competetion**

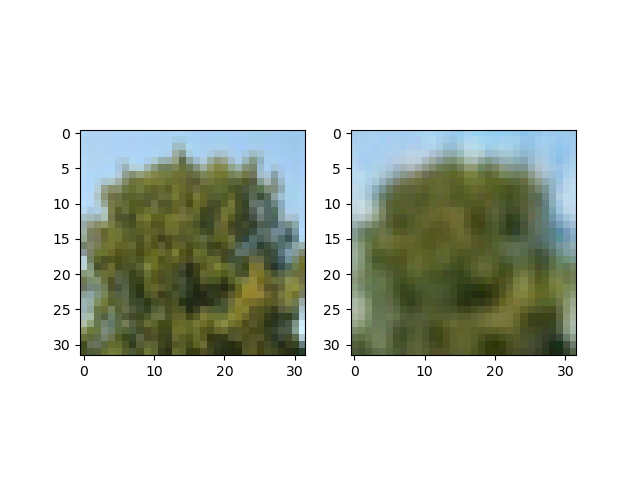
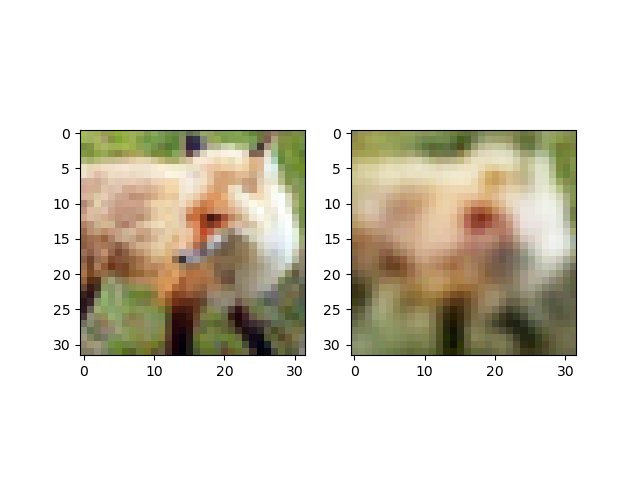
1. (1%) 請附上你在kaggle競賽上表現最好的降維以及分群方式，並條列五種不同降維維度的設定對應到的表現(public / private accuracy)，auto-encoder 和 PCA 只要任一維度不一樣即可算是一種組合。

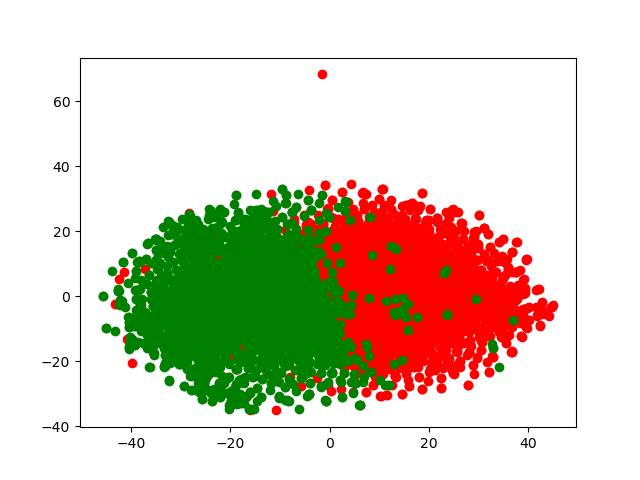
我在kaggle中表現最好的降維方式和分群方式為 auto-encoder 搭配 PCA(1024→100)+TSNE(100→2)+K-means。  
因為一開始準確率一直上不去，卡在79%左右，所以去查了降維方式的相關資料，發現TSNE的表現似乎跟數入的維度有關，因此作了以下的測試。

下表列出了先用PCA降到不同的維度，再用TSNE降到2維的表現。可看出先降到抬高或太低維度表現都不是最好的，個人實驗結果是降到100維表現皆最好。

|  |  |  |
| --- | --- | --- |
| PCA維度 | Private | Public |
| 50 | 0.78355 | 0.77600 |
| 100 | 0.81488 | 0.80866 |
| 150 | 0.77133 | 0.76844 |
| 200 | 0.69977 | 0.68466 |
| 500 | 0.49488 | 0.49822 |

1. (1%) 從 kaggle 的 dataset 選出 2 張圖，並貼上原圖以及經過 autoencoder 後 reconstruct 的圖片；請將 visualization.npy 的檔案降維至二維平面並利用給定的 label 將資料上色 ( 前一半為 0；後半為 1 )。

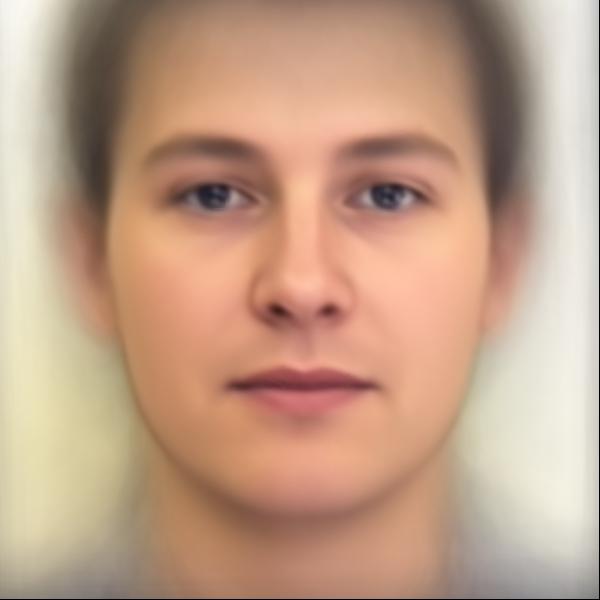




**Eigenface**

1. (1%) 請畫出所有臉的平均以及 Eigenvalue 最大的前五個 Eigenfaces。

有嘗試過以float32和float做計算，以下皆呈現float64做計算的結果(較好)。



臉的平均

前五大 Eigen Value 對應到的 Eigen Face 分別如下

Top1 Top2

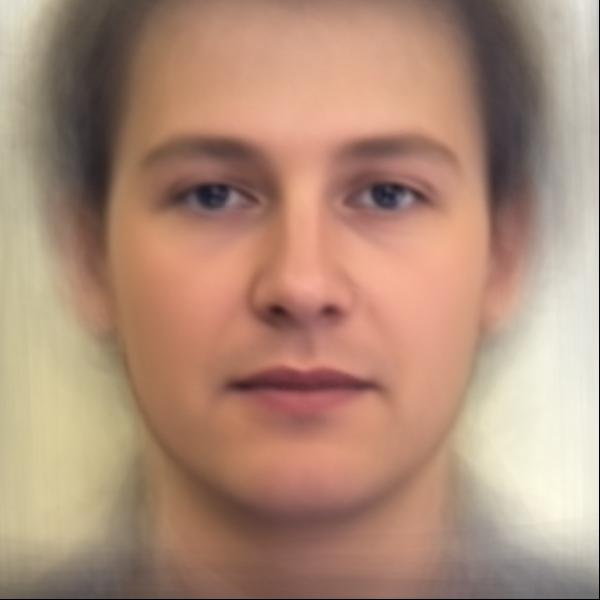
 

Top3 Top4

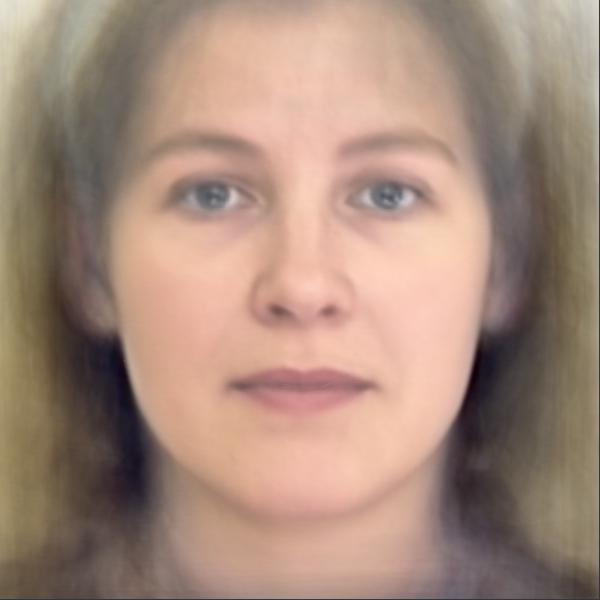


Top5

1. (1%) 請從數據集中挑出任意五張圖片，並用上題前五大 Eigenfaces 進行 reconstruction，並畫出結果。

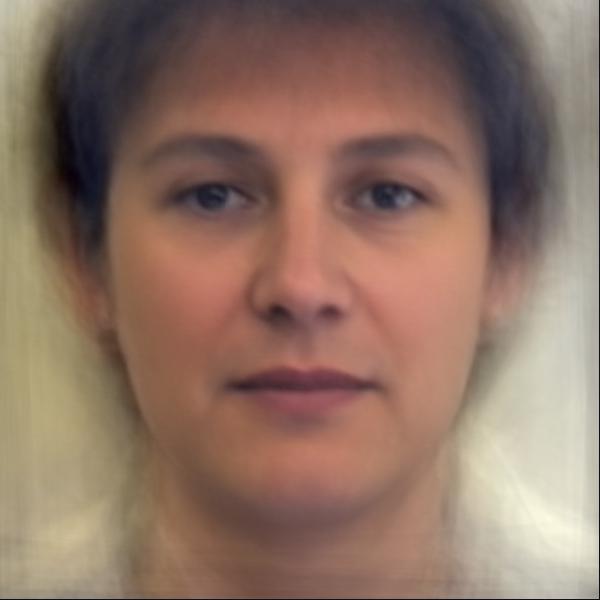
7.jpg & reconstruction

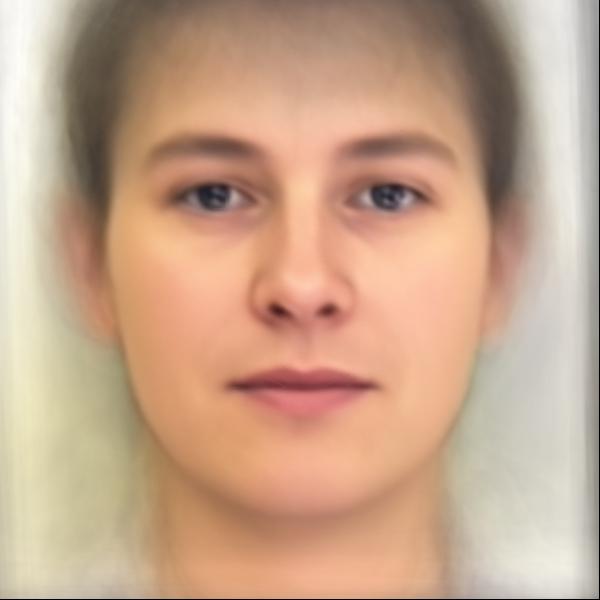
119.jpg & reconstruction

138.jpg & reconstruction

157.jpg & reconstruction

133.jpg & reconstruction

1. (4%) Refer to math problem：  
   <https://hackmd.io/@g4HRMJCzQL2hzLedRcbVPQ/SyCBoc1qt>