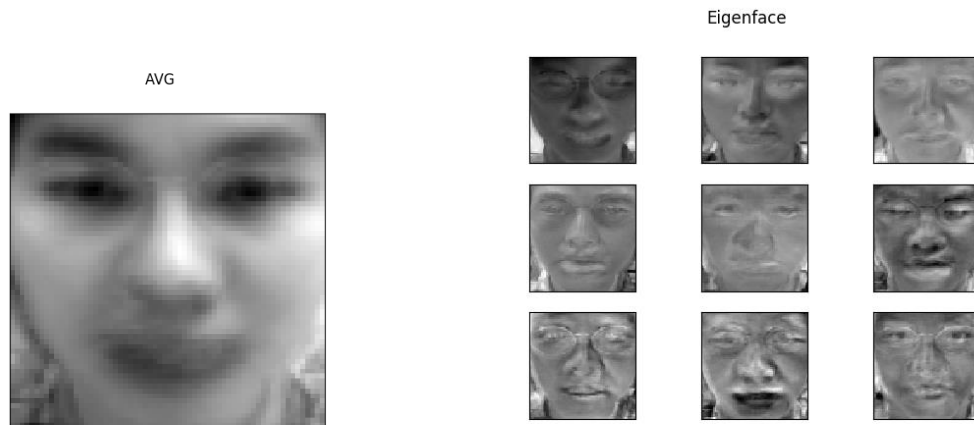


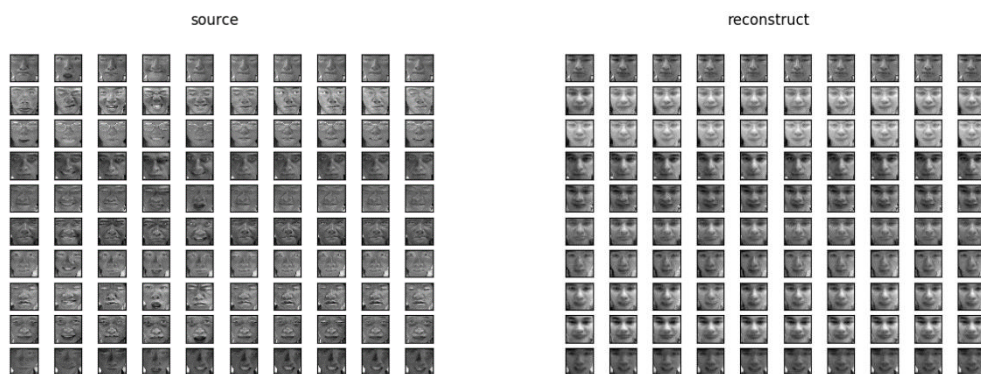
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

答：(左圖平均臉，右圖為 3x3 格狀 eigenfaces, 順序為 左到右再上到下)



1.2. Dataset 中前 10 個人的前 10 張照片的原始圖片和 reconstruct 圖 (用前 5 個 eigenfaces):

答：(左右各為 10x10 格狀的圖, 順序一樣是左到右再上到下)



1.3. Dataset 中前 10 個人的前 10 張照片投影到 top k eigenfaces 時就可以達到  $< 1\%$  的 reconstruction error.

答：(回答 k 是多少)

k=60

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

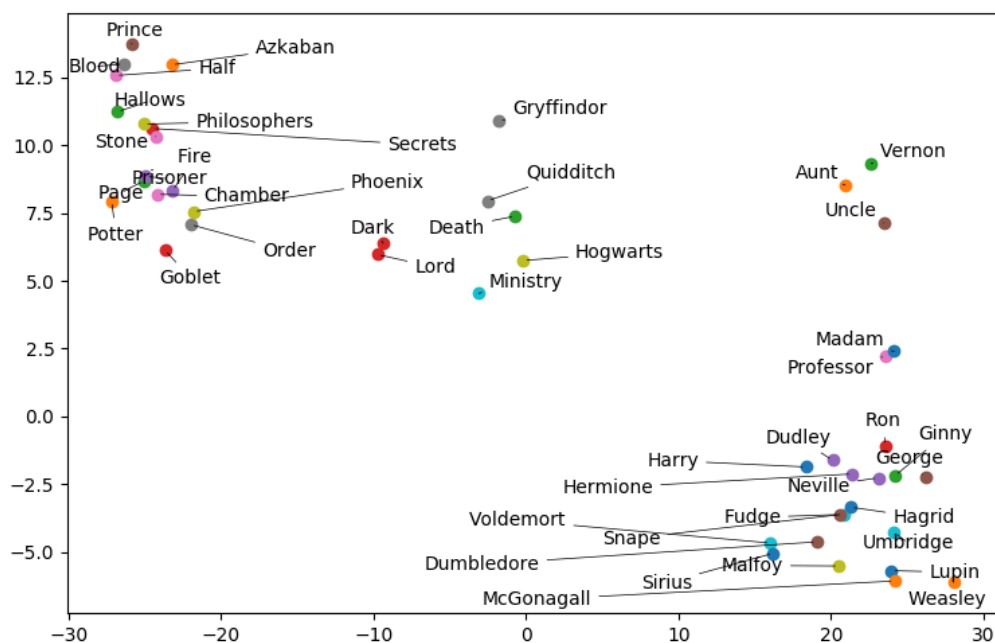
答：

Size: 這個是指維度的大小

Min\_count: 這個是指只有當超過這個頻率的單字才會被拿進來 train。

2.2. 將 word2vec 的結果投影到 2 維的圖:

答：(圖)



2.3. 從上題視覺化的圖中觀察到了什麼？

答：

我看到了和混血王子有關的詞都聚在一起，然後混血王子附近還有一群死神聖物相關的詞彙，然後和哈利和他朋友們會比較容易出現在一起。然後他姑姑和叔叔和他們的姓混在一起，我們可以從這個圖中觀察出哪些詞彙之間的相關性比較高，也能從這之中了解人物之間的一些關聯性。

3.1. 請詳加解釋你估計原始維度的原理、合理性，這方法的通用性如何？

答：

我的做法是拿原本產生測資的程式先 generate 出一些原始維度對應到 100 維度的資料後拿去算出 eigen vector，然後把那些相同原始維度所算出來的 eigen vector 的結果取平均，最後要算答案的時候就拿 test data 所算出來的 eigen vector 去看和之前每個原始維度所算出來的 eigen vector 的 RMSE 差最小的就當成答案，而這樣的做法在 kaggle 上拿到了 0.09576 的成績。

我覺得這個方法的合理性很高，因為他所出來的答案會和原始資料會有很高關聯性，因此這個做法很合理。

而這個做法在通用性上的缺點就是如果沒有辦法取得原本產生測資的程式的話，就沒有辦法使用這個程式了。

3.2. 將你的方法做在 hand rotation sequence dataset 上得到什麼結果？合理嗎？請討論之。

答：

如果把我們把圖片先壓成 10\*10 之後丟到 3.1 所 train 出來的 model 中，會得到他大約接近 16 左右的維度，但是因為壓成 10\*10 之後可能會丟失太多的資訊，因此我後來把圖片壓成 48\*48 的大小後把 eigen vector 拿出來看，會發現比較大的項的數量也大約接近 15~16 這個數字，因此我們推測這張圖的降維之後大約為 15~16 維左右。

而我覺得這件事情還滿合理的，因為我們會發現圖中手和圓盤的 U 形邊緣大約有 7~8 個(五個手指和圓盤的兩端)，而每個 U 形邊緣會需要花約兩個維度去描述，因此得到大約 15~16 維附近的數字是滿合理的。