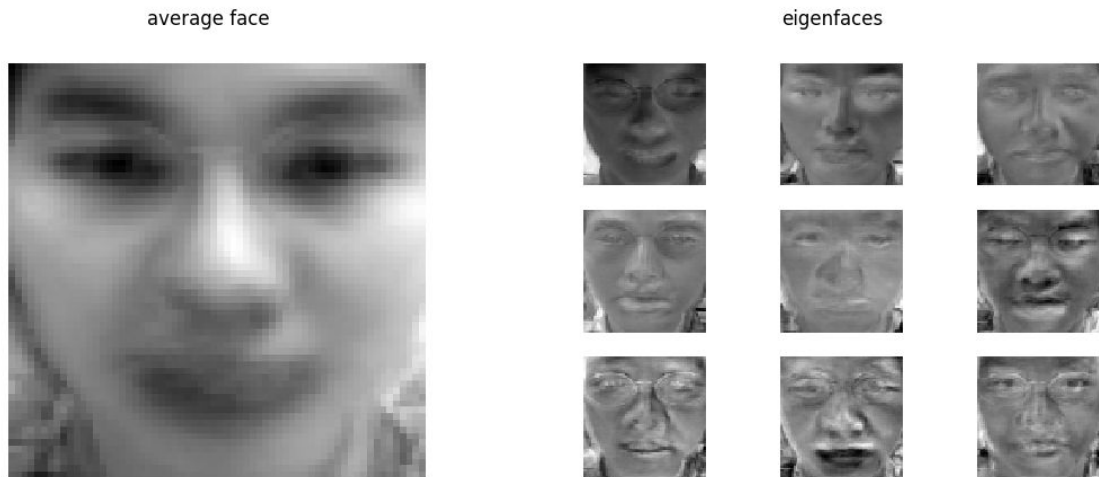


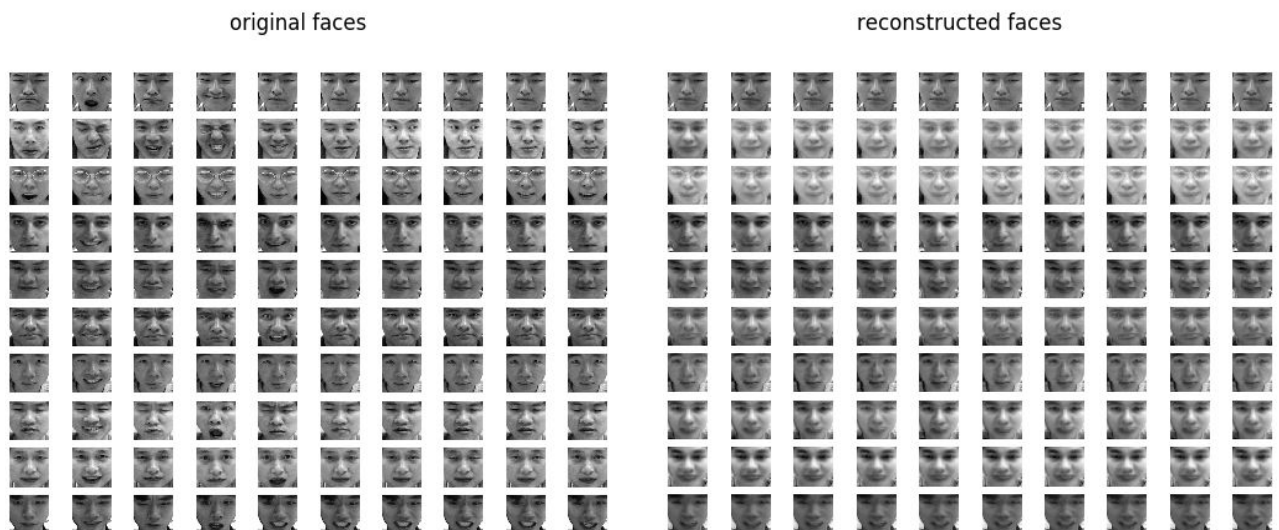
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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:

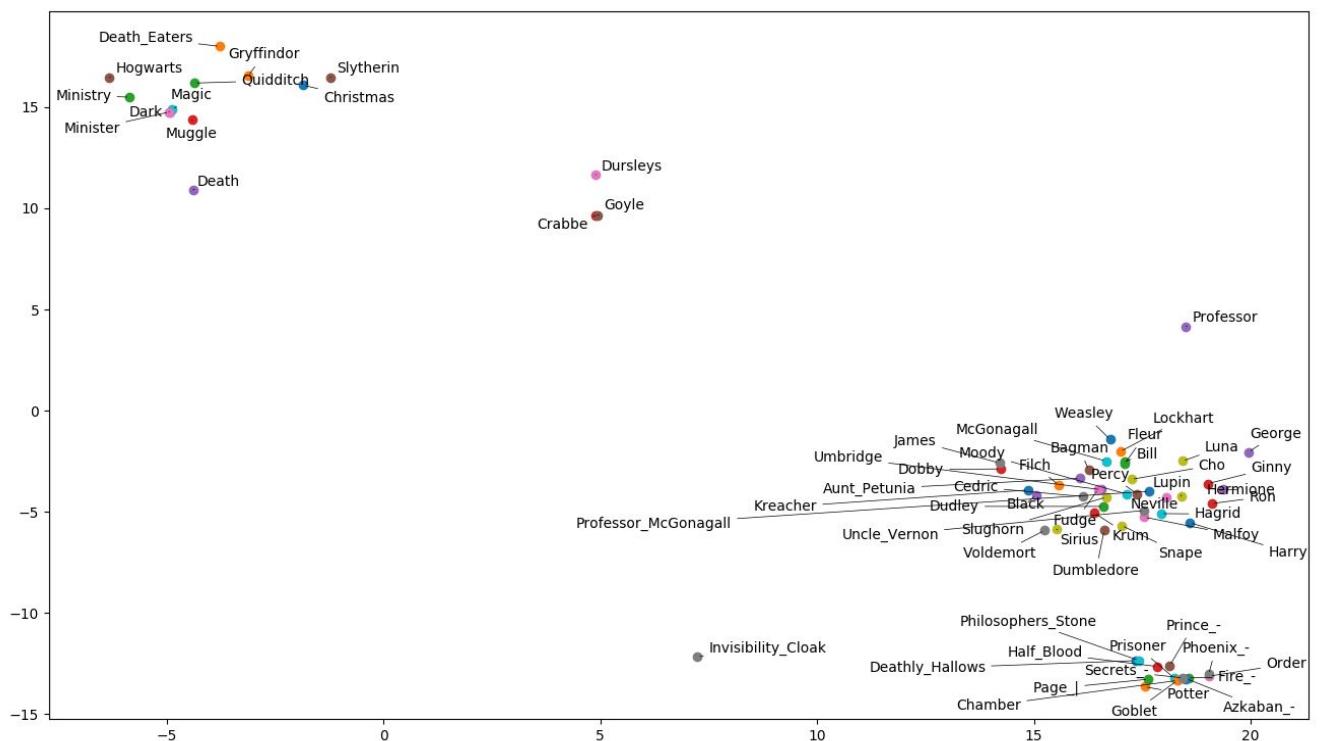
答：

- cbow: 使用 CBOW 或是 Skip-gram 去 train
- window: 一個 context 要取多少 word
- sample: downsampling

我使用 CBOW, window 設 5, sample 設 1e-3。

## 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：

主要觀察到五點：

1. 跟書名有關的在一起
2. 人名都在一起，除了哈利原本在的麻瓜家庭 Dursleys
3. 魔法與麻瓜、黑暗以及魔法部都很相近
4. 葛來分多與史萊哲林以及魁地奇都很相近
5. 隱形斗篷特立獨行

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：

主要的基本想法如下：

1. 模擬原本 data 可能產生的方式去自己生出 training data
  - x: 利用 ball tree 找出 nearest neighbors 再算出 average eigenvalues
  - y: dimension
2. 用 linear SVR 跟自己產生的 training data 去 train
3. 使用 train 好的 model 去 predict testing data

在知道 training data 真正的 dimension 或是可以有效估計的情況下，這方法是可行的，但一般只有 testing data，無法像這題合理生出正確的 training data，因此這個方法的通用性可能不高。

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：

我使用 PCA 來做降維的動作，並找出 variance ratio 加起來超過 0.9 (90%) 的維度來估計真實維度。

做出來是 23 維。

但不太合理，因為 reconstruct 出來竟然是錯的。

