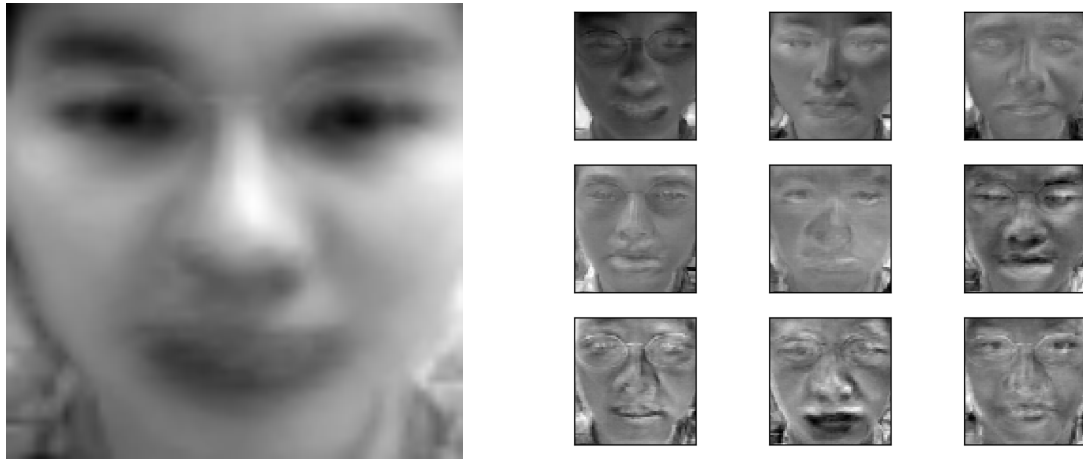


1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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)



可發現地前五個 eigefaces 並不夠，在表情上很多圖片還是有點差異

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：k 為 60

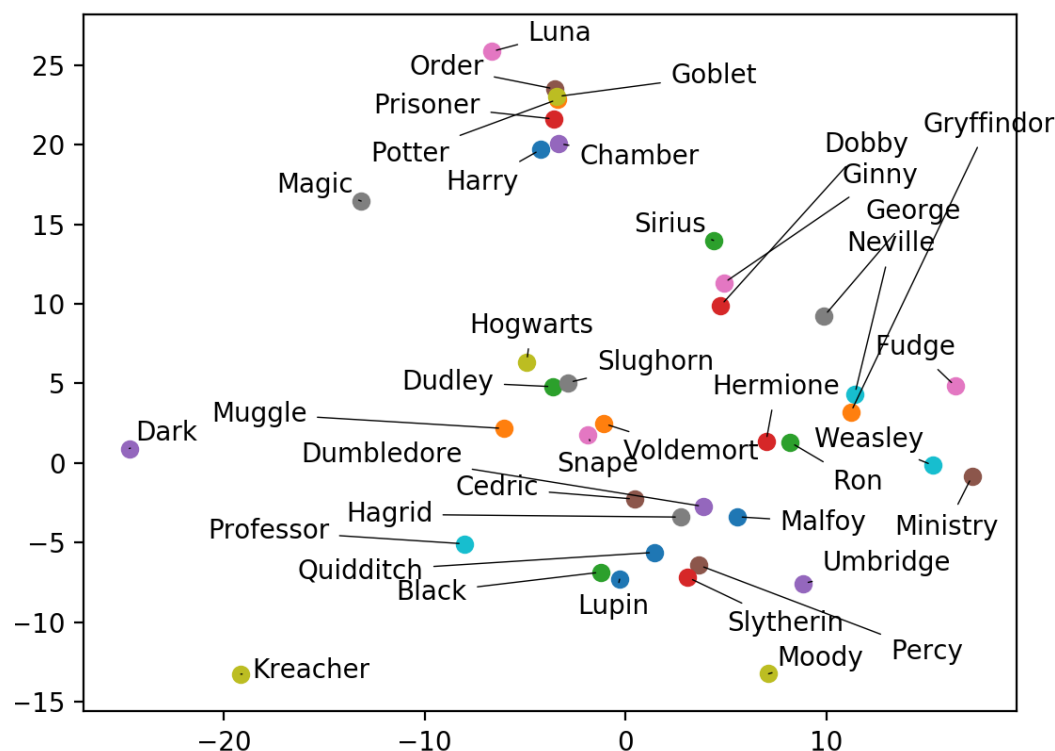
```
smallest 60
chenyu:hw4 peter$
```

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：CBOW=1 代表使用 CBOW 演算法，size 代表每個詞要用幾維的向量表示，negative 表示負採樣參數，每有一个正樣本，選擇幾個負樣本的參數。min-count 為控制詞典大小的參數，可以標出高頻詞，因若這個詞出現的次數小於 min_count，那他就不會被視為訓練對象。而 window 代表在 train 的過程會取到上下文的幾個詞。下圖為給訂的參數

```
MIN_COUNT = 6
WORDVEC_DIM = 300
WINDOW = 5
NEGATIVE_SAMPLES = 10
ITERATIONS = 5
MODEL = 1
LEARNING_RATE = 1e-3
```

2.2. 將 word2vec 的結果投影到 2 維的圖:



2.3. 從上題視覺化的圖中觀察到了什麼？

答：可以發現 harry 跟 potter 相近，ron 與 weasley 也相近，代表他們其實都具有同樣的語義（即為同一個人，而 ron 跟 hermione 相近也很合理，在劇情中榮恩跟妙麗本來便有很大的相關性

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：

估計方式為對不同維度設定不同的 threshold，用助教的 code 生成 data，並去分析不同維度下的 threshold 為多少，以此方法來判斷資料的維度，而這方法應是非常合理且相當通用，因不同 dimension 的 data 所用的 threshold 本來就應不同，缺點為必須有 data 分析不同 dimension data 的 threshold。

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：將我的方法套用在 `hand rotation sequence dataset` 上，我只取前 100 張相片，當作 100 維的 data，用 pca 降到 60 維之後，再用前一題的 threshold 去判斷維度，答案結果跑出來是 4 維，我認為是一個合理的答案。因此這個方法應該是蠻通用且合理的。