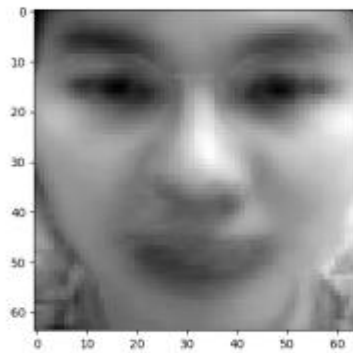
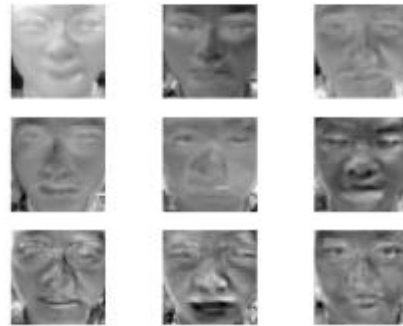


1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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



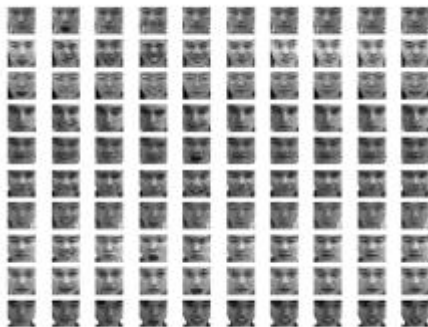
圖(一) 平均臉



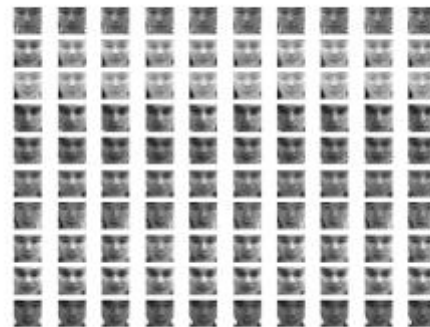
圖(二) 3X3 eigenfaces

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)



圖(三) 原始圖片



圖(四) reconstruct 圖

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 = 59

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：

min\_count = 50 : 過濾掉出現次數低於 min\_count 的詞。

size = 100 : 把每個詞投影成 100 維的向量。

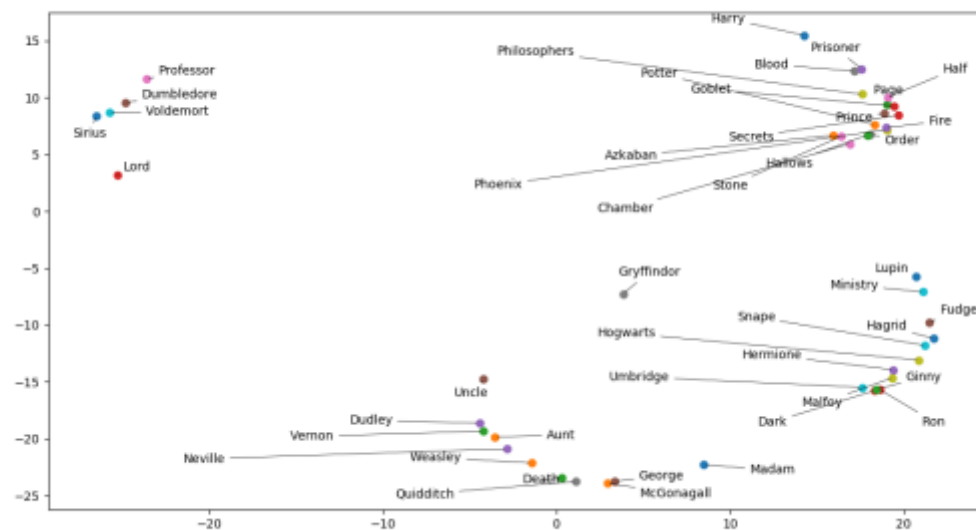
**cbow=1:** 使用 CBOW 模型，用詞  $w$  的上下文去預測  $w$ ，而另外一個常用 Skip-gram 模型是用詞  $w$  來預測  $w$  的上下文。這兩個模型都是用 **gradient descent** 的概念取代傳統類神經網路的訓練方式，大大減少複雜度。

**negative = 10:**  $>0$  表示使用 **negative sampling**。在 CBOW 模型中，已知詞  $w$  的上下文  $\text{Context}(w)$  去預測  $w$ ，因此對一個  $\text{Context}(w)$  來說，詞  $w$  就是一個正樣本，而其他詞就是負樣本。原本每次更新，我們會對所有正樣本和負樣本更新，目的是增加我看到正樣本的機率同時，降低沒看到某詞(負樣本)的機率。做 **Negative sampling** 目的是每次會隨機抽取幾個點(我選 10 個點)更新，而不會對所有負樣本更新，以降低複雜度。

**Window = 5:** 輸入上下文的大小，若窗格太大會訓練到一些沒關係的詞，所以我設 5  
**Iter\_ = 10:** epoch 遞迴次數

## 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：

可以看到整張圖大概分成 4 個群落，關係近的点會聚在一起，可以觀察到 Uncle、Dudley、Aunt、Vernon 這些詞靠的很近，而這些詞的確關係密切，他們是哈利的麻瓜親戚，Vernon 是哈利叔叔的名字，而 Dudley 是哈利的表弟。而可以看到 Harry 和 Potter 這兩個詞都在右上角，和 Philosophers、Prisoner、Azkaban、Goblet、Fire、Prince...關係很近，可以發現這些詞都是哈利波竹特每集小說的名稱：Harry Potter and the Philosophers Stone、Harry Potter and the Prisoner of Azkaban、Harry Potter and the Goblet of Fire 等。在左上角有 Professor Dumbledore(鄧不利多)、Sirius(天狼星)、Lord Voldmort(佛地魔)，這三個詞都是《哈利波特》著名的角色，從三個詞靠的很近可以推測，在小說中這三位有非常多的互動。

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：

先產生額外的 **data sets** 去訓練，每個 **data set** 隨機取樣 **N** 個點，每個點找出離他最近的 **k** 個點後去計算這個 **subset** 的 **eigenvalues**，最後再把這 **N** 個點的 **eigenvalues** 平均後用 **SVR** 去找出最 **fit** 資料分佈的 **hyperplane**(維度)。再用這個 **hyperplane** 去預測原始維度。

但這個方法通用性蠻低的，因為這個作法我們必須先知道原始維度的範圍，才能去產生額外的 **data sets** 去訓練，但其實在大部份的情況我們是無法得知原始維度的就不能使用這個作法，因為原始維度就是我們想知道的答案。

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：