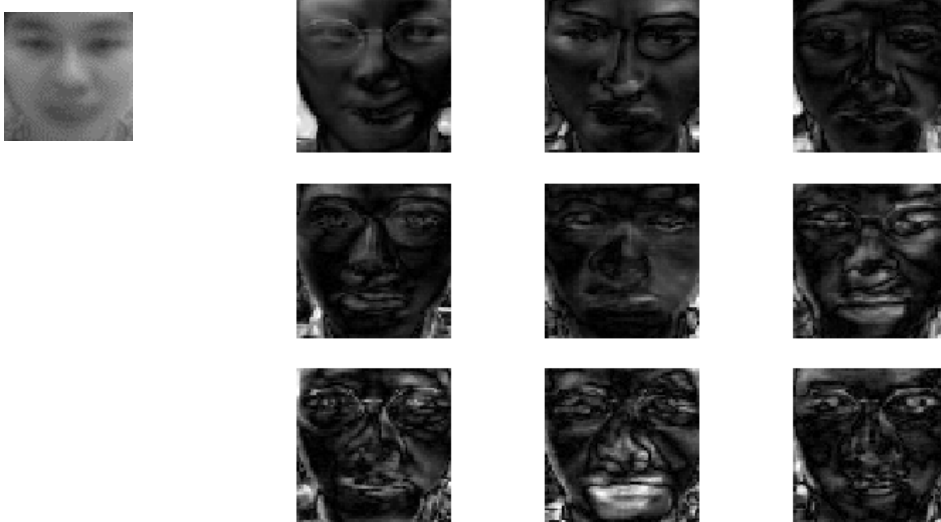


學號: R05942114 系級: 電信一 姓名: 方敬勻

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)

90

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答:

Train : train data

Cbow : 使用的 model (1)

Size : 書出 vector 的 dimensions(150)

Min\_count : word 被加入 train data 的最小次數(4)

Window : 前後文取樣的範圍(5)

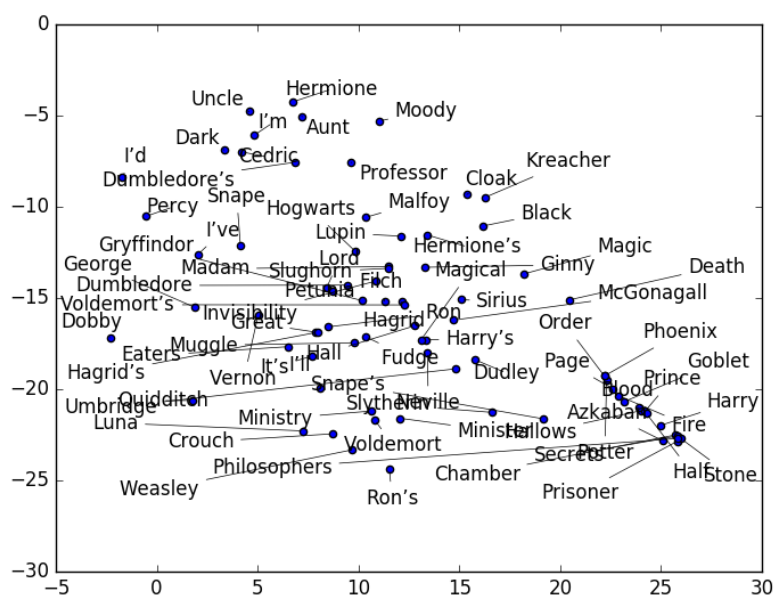
Sample : 高頻率詞的採樣率( $1e-5$ )

Iter : 迭代次數(5)

Alpha : learning rate(0.025)

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:

從圖的最右邊可以看到 train 出來的結果有 half, blood, prince 等詞與可以知道這邊所

Train 出來的關係很有可能都是從混血王子背叛那邊所找出來有相關的詞語，看到中間 voldemort, crouch, Hall 這些東西都在靠近的地點可以假設出 word2vec 所找出來的 wordvector 可能在小說中從在一些劇情的連接的關係。

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：

因為知到 data 產生的過程所以可以自己產生 data 後,就可以借由隨機產生的 data 利用相近的點得出 model 找出可能的 dimension

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：