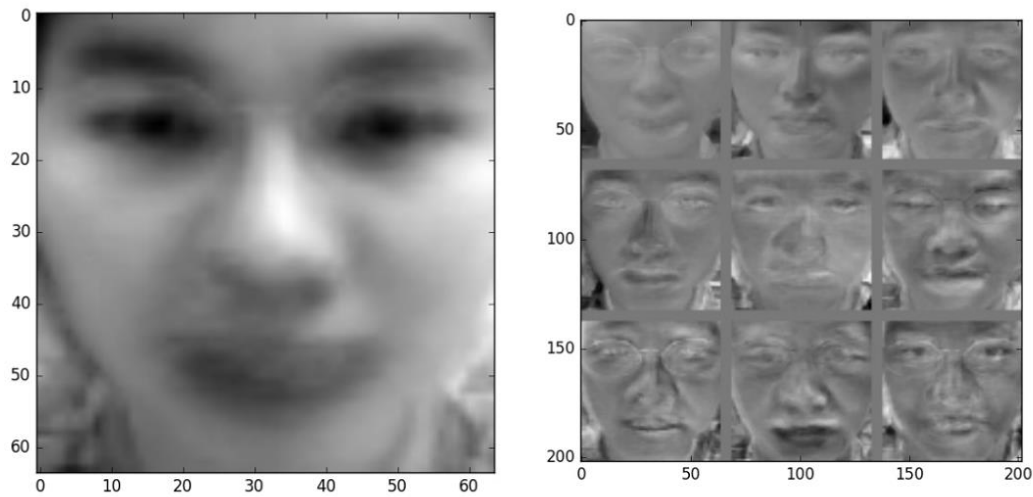


學號：R05943129 系級：電子碩一 姓名：侯人文

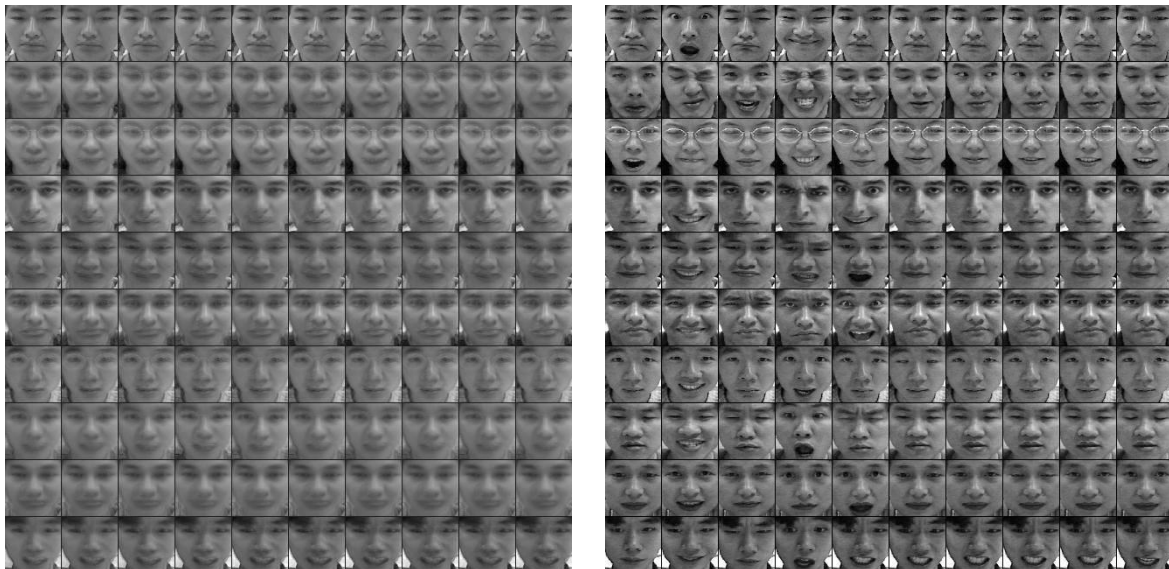
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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)

65

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：

Wordvec.word2phrase

使用 word2phrase 輸出訓練模型。

Wordvec.word2vec

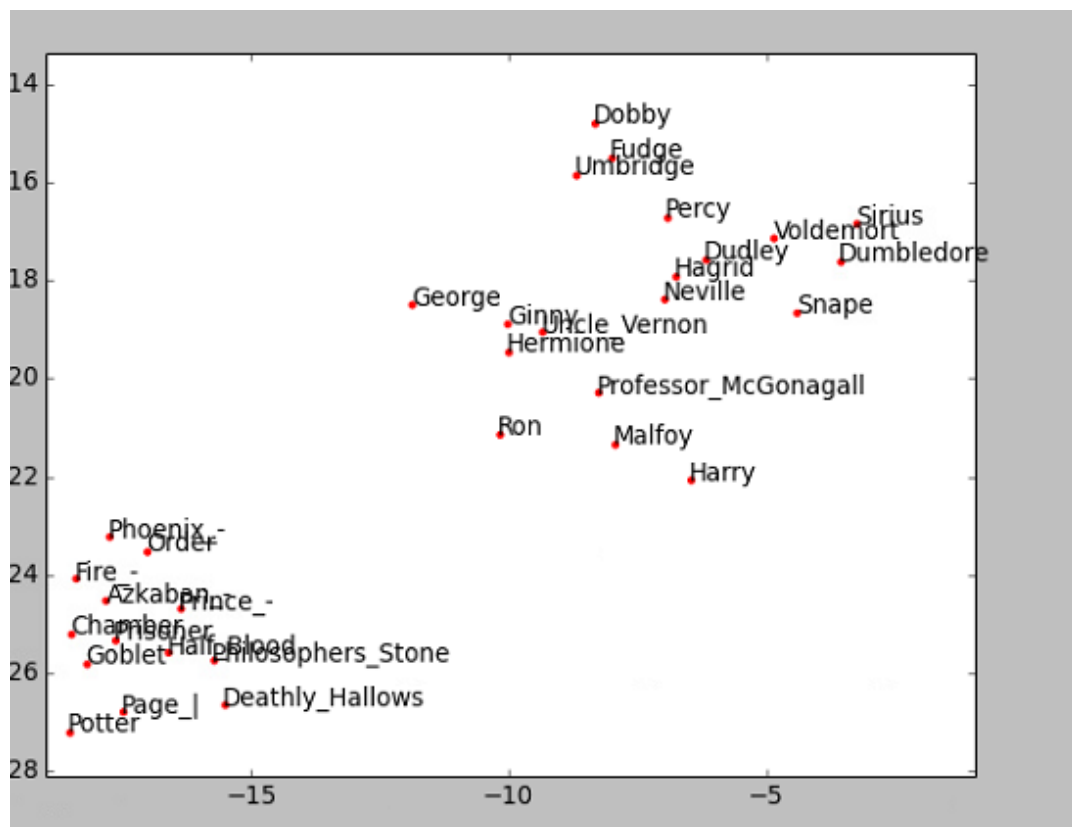
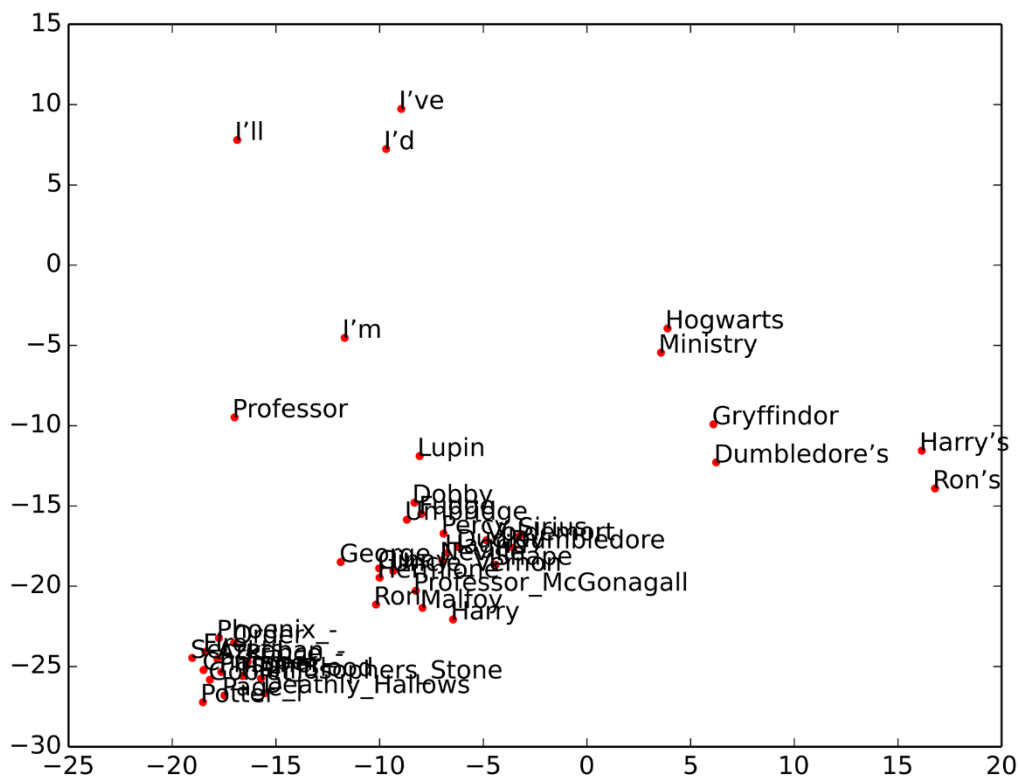
它生成一個包含二進制格式的字向量的 XX.bin 的文件。

Wordvec.load

把 XX.bin 的資料給 load 進來

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：我看到其實他會把 I'm 、I'll 、 I'd ，這些和其他的字彙分得很開，下圖是上圖放大的，可以看到他會把人名集中在一起，但不知道為甚麼 Harry 和 Potter 分得蠻開的，感覺就是比較常叫 Harry 所以這兩個字不會靠得太近，還可以發現 Herry's 和 Ron's 靠得蠻近的，應該這兩字蠻相近的。

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：

我使用這篇論文 [Maximum Likelihood Estimation of Intrinsic Dimension](#) 上所提供的方法來估計這次的作業，一開始我直接使用 PCA 來估計，但是感覺是因為是線性的關係，kaggle 上最好只能到 0.19 多，後來我使用這篇論文技巧，一開始他在先對搜尋附近範圍的維度，在測出歐幾里得距離來計算出完整矩陣的距離，最後估計維度，最後其實產生出來的結果大概 0.16 多，之後我把許多可能超出的維度隨便設定一個值給他，在 kaggle 上可以跑到 0.1 左右

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：原本我是拿 20 點來估計，得到大概 200 個 dim，但在 rotation 後回得到大概 400 個 dim，其實我不是很了解他的概念大概是什麼，產生出來的結果也不是很合理。