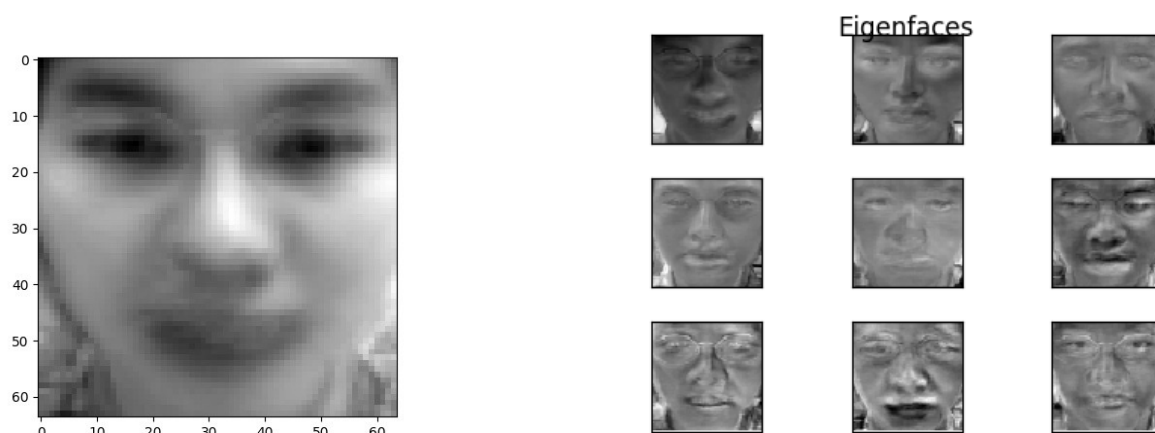


1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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 = 94

reconstruction error = 0.00971929783795

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答:

size(word vector 的尺寸) = 100

min_count(出現少於此次數的 word 會被忽略) = 5

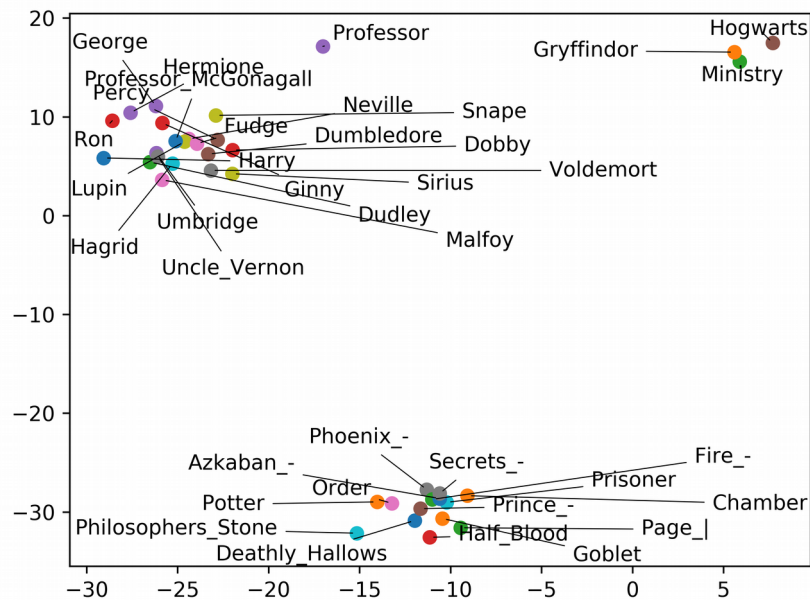
window(word 之間能跳過的最長距離) = 5

negative(negative examples 的數量) = 10

$\alpha(\text{learning rate 初始值}) = 0.025$

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:

我們可以看到大概可以將幾個不同類別的單詞做一個分類，像是書名的字詞大多出現在下面，右上角是學院跟學校的相關字詞，學生跟教授的名字則被分在左上角，所以是有些分類的效果出來，只是分類還不是很精細而已。

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:

方法:

1. 利用 `gen.py` 生出原始維度 1~60 的 `labeled_data`，作為 `training data`
2. 利用 `NearestNeighbors`、`SVD` 求出 `average eigenvalues`
3. 利用 `SVR` 找 `average eigenvalues` 跟原始維度的回歸模型
4. 調整 `NearestNeighbors` 和 `SVR` 的參數

原理:

原先的 `oracle network` 與 `SVD` 有些類似，而增維的過程中應該可以找到維度間有相似的 `NearestNeighbors`，因此利用這兩個方法找出 `average eigenvalues` 後做 `SVR` 推測。

通用性:

1. `SVR` 假設是線性解
2. 受亂數產生的 `training data` 影響很大

3.Domain knowledge 改變準確性會降低

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：

得到負數的不合理結果。

原因大致上如上題所說的通用性問題，hand rotation sequence dataset 不符合原本的 domain knowledge，因此得出來的結果便不如預期。