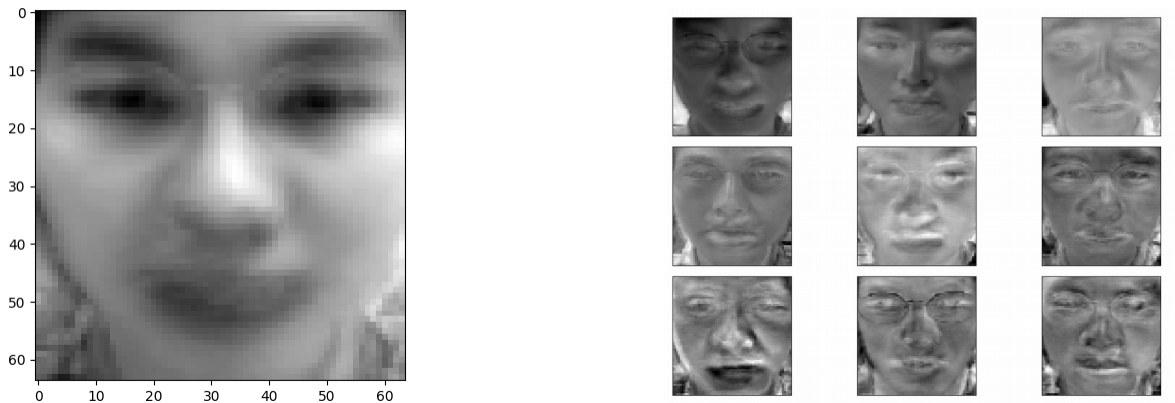


學號: B05901189 系級: 電機一 姓名: 吳祥勸

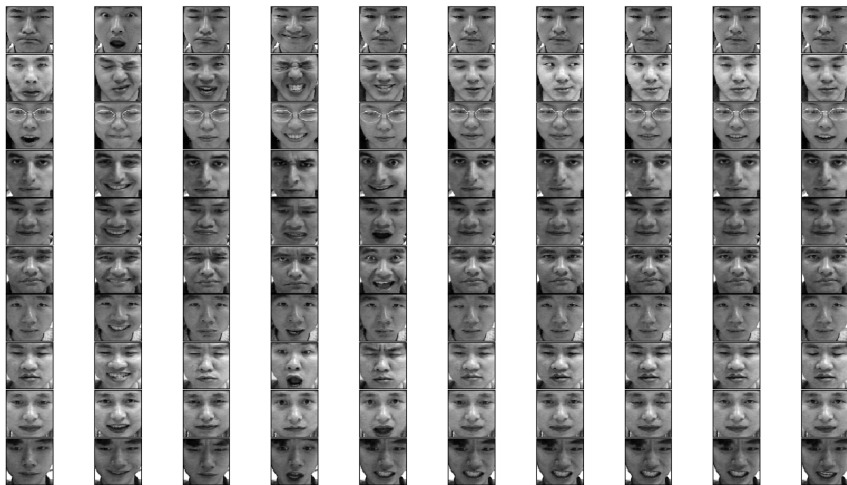
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

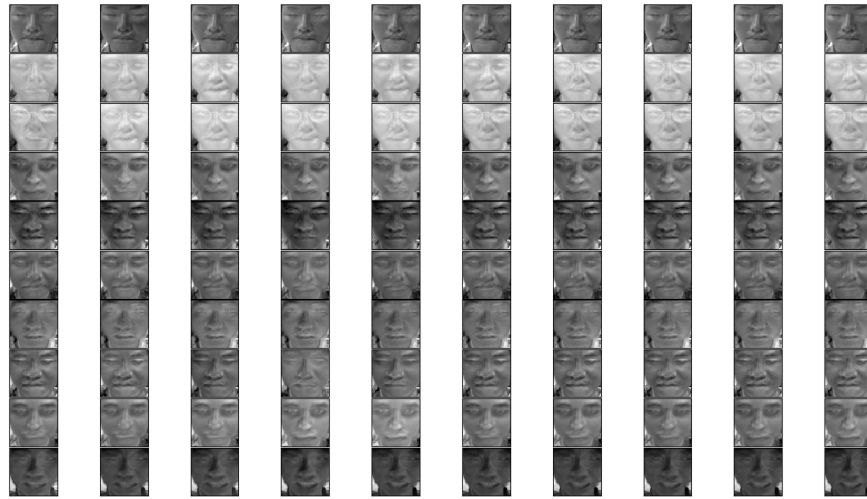
答: (左圖平均臉, 右圖為 3x3 格狀 eigenfaces, 順序為 左到右再上到下)



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)



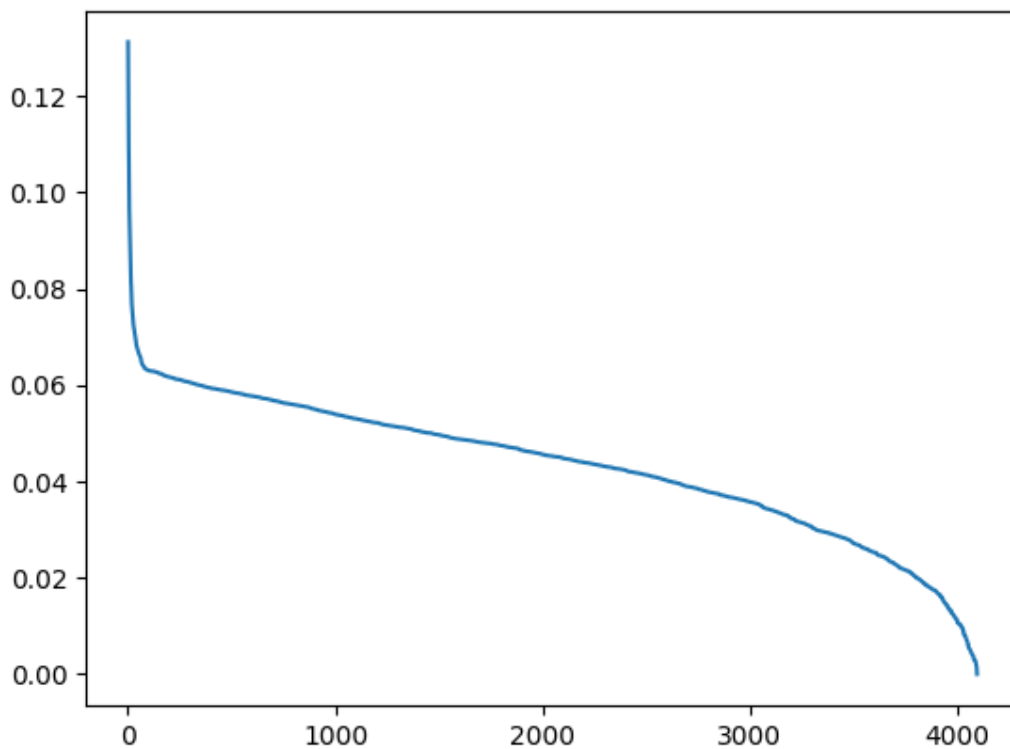


1.3. Dataset 中前 10 個人的前 10 張照片投影到 top k eigenfaces 時就可以達到 $< 1\%$ 的 reconstruction error.

答: (回答 k 是多少)

k=4020

reconstruction error—k 關係圖



2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答:

`word2vec.word2vec("all.txt", 'all.bin', size=100, verbose=True)`

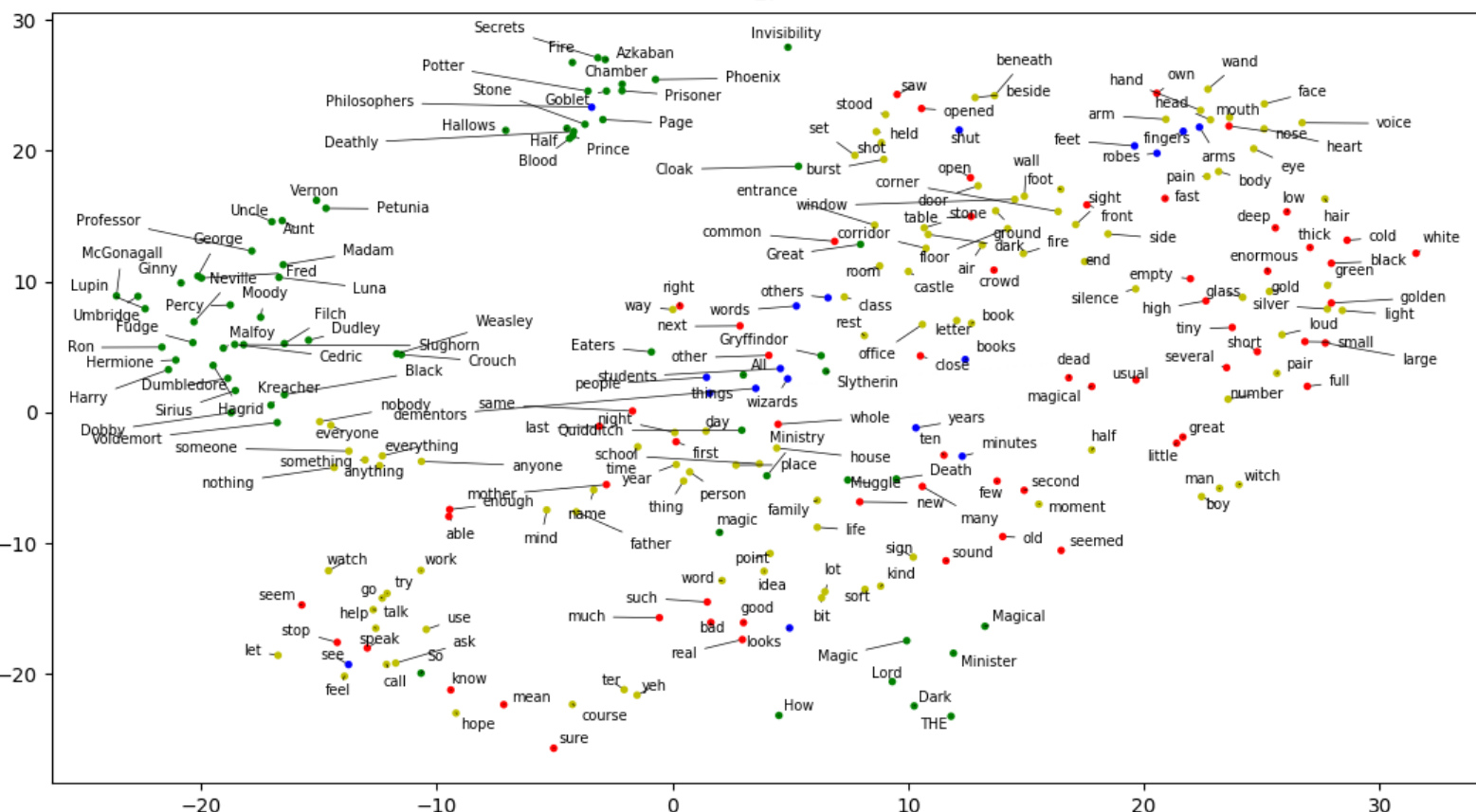
用 all.txt 訓練

model 存成 all.bin

把每個字轉成 100 維的向量

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：

相同性質的字會在附近，比如左邊綠色那塊都是人名，左下角都是動詞，右上角都身體部位。

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：

將一整組 data 做 pca。把 explained_variance_ratio 拿去 train 一個 DNN。

我沒有像助教那樣找附近的點再去做 pca，所以這個方法應該只適用在沒有太彎曲的高為平面，因為原本資料產生的方式上參數都是 mean=0 variance=1 的 normal distribution 取出的，數值都不會太大，剛好符合這樣的特性。

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：

5.4213233

理論上照片應該都是二維平面，所以這個估計太大了。不過原本產生 training data 的方法是一個 neural network，和 hand rotation dataset 的性質也不完全相同，所以我覺得要做到更接近 2 可能必須找不一樣的 dataset 來訓練。