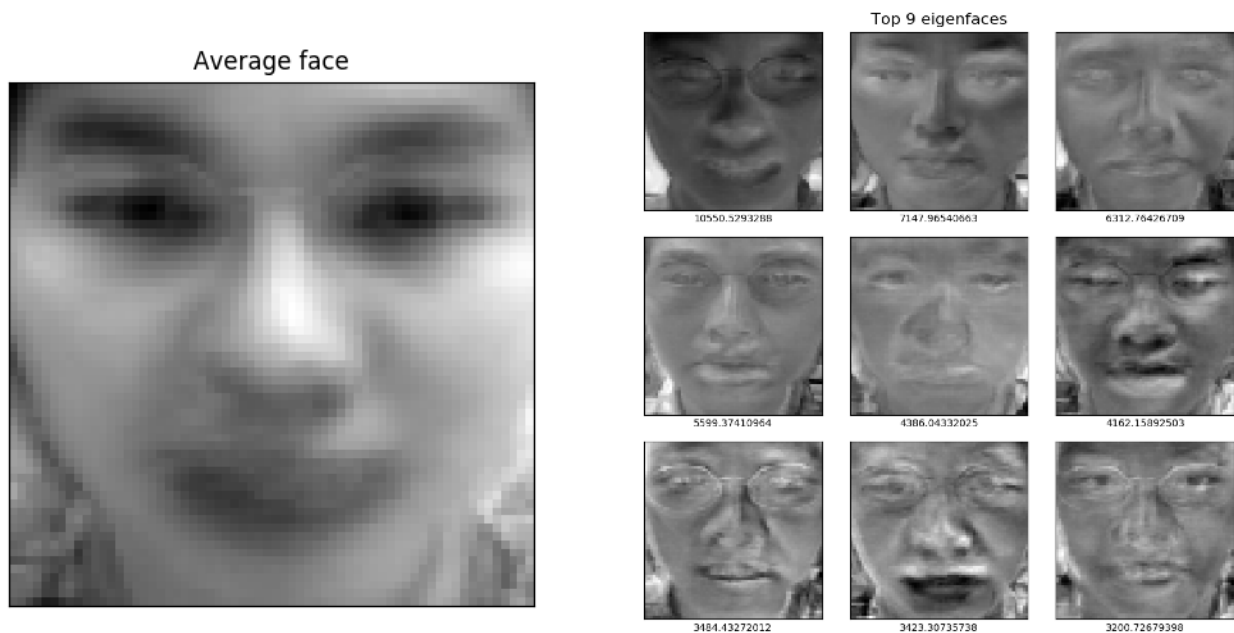


學號：B03901039 系級：電機三 姓名：童寬

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，圖的下標為 eigenvalue)



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 = 59

rmse = 0.9968%

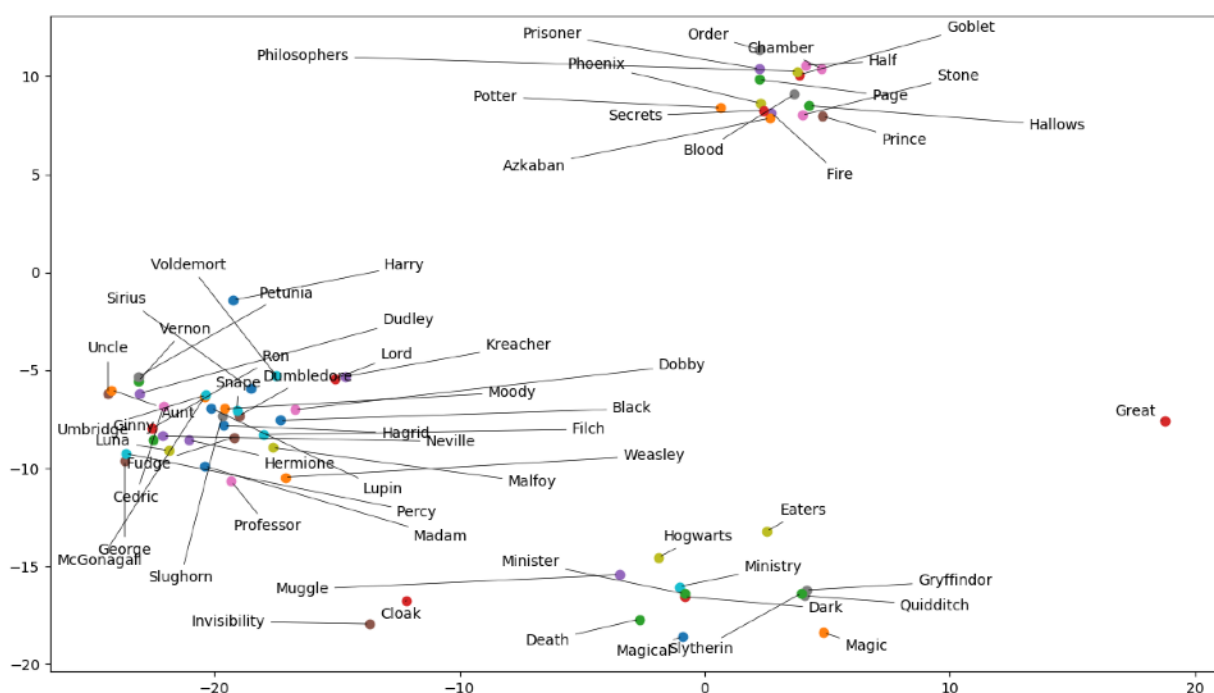
### 2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：

參數	初始值	我使用的數值	意義
size	100	100	每個 word vector 的 dimension
window	5	5	決定多長的前後文算是有關係的 context
sample	'1e-3'	'1e-3'	出現頻率比此值高的字會被 down sample
hs	0	0	決定是否使用 Hierarchical Softmax 指在算 softmax 前先判斷輸出的字是哪一類的 (要事先做分類)，再判斷是那一個字
negative	5	5	選擇幾個 negative samples 指的是跟目標字沒有關係的 sample context 它們在目標函數中要被最小化
threads	12	12	看要用多少 threads 去計算，可以加速訓練時間
iter_	5	5	看要訓練幾個 iterations
min_count	5	5	出現次數少於此值的字，不會列入訓練
alpha	0.025	0.025	起始學習速率
debug	2	2	debug 的模式，2 表示有較多的 debug 訊息
binary	1	1	決定是否將 word vector 以 binary 方式儲存
cbow	1	1	決定是否使用 continuous back of words model 指的是給周圍的字去算目標字的機率
save_vocab	None	None	決定是否儲存 vocabulary
read_vocab	None	None	決定是否讀取 vocabulary
verbose	FALSE	TRUE	決定是否印出 output

## 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：

- 可以大略分成三個區塊，右上、左下跟中下。
- 右上的那群字幾乎都是每一集的書名裡的字。
- 左下則是幾乎都是人名，first 跟 last name 都出現在這裡。
- 中下則是比較多專有名詞，像是 Hogwarts, Gryffindor, Slytherin 等。
- 從三大區塊的字詞組成，的確可以看出相似的詞有分在一起。有些比較小的例子也可以看出，像是 aunt 跟 uncle 兩個幾乎重疊，Gryffindor 跟 Slytherin 也是幾乎重疊。

### 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：

原理（合理性）：

- 在一組 dataset 裡，可以透過找一小部分 data 的分布去預測整體 data 的分布。
- Eigenvalues 可以代表對應的 eigenvectors 在 dataset 中所佔的份量。可以透過取不同數量的 eigenvectors 去當 dataset 的基底向量，也就是用較少維度去描述該 dataset。在這個問題中，可以想成是有多少夠大的 eigenvalues 就代表應該有多少 eigenvectors 當基底向量，也就代表原始維度的大小，因此平均 eigenvalues 是跟原始維度相關的。

方法：

1. 隨機以助教提供的程式生成原始維度為 1~60 的 data。
2. 每個維度生成五組 dataset，分別是資料數為 10000, 20000, 50000, 80000, 100000 的 dataset。
3. 隨機從每個 dataset sample 50 點。
4. 找出每一點的 200 個 nearest neighbors，並把每一組分成一類。
5. 算出每一組的 eigenvalues，並把算出來的值除以每一組最大的 eigenvalue。
6. 算出每一組 eigenvalues 的平均。
7. 以 SVM 做 supervised 的訓練，求出給定平均 eigenvalue 能輸出 dimension 的 model。
8. 用此 model 預測原始維度（因為維度都是整數的，所以有對結果做 rounding）。

通用性：

- 因為一開始產生 data 有用到額外的資訊（ELU, weight 跟 bias 的大小範圍等），因此這個方法不太能廣泛的應用。如果要用在其他問題上，一定要在一開始產生 data 那裡做一些更動。

### 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：

因為這兩題的 data 很不一樣，因此我有在一開始產生 data 那裡做一些改變：

1. 照片太大 (512×480)，我把它們 resize 成 64×64
2. 原本的 intermediate 層的維度從 60~80 改成 100~3000
3. 每個維度只產生一組 dataset，資料數為 481 (hand rotation 照片的數量)

結果：

- 原始維度 = 1，但因為我有把小於 1 的值直接當成 1，所以其實 SVR 直接出來的值是 -1。
- 這樣的結果可能代表的是原始維度真的很小。這也還算合理，因為這些照片就只是手握著杯子旋轉而已，可以用 360° 去描述他們。
- 但從理論上來看，這個方法是不合理的，因為我在產生資料的時候，用了很多無關的假設，例如：ELU, weights, biases 等。我覺得用第一題第三小題的方法可能會更具合理性。