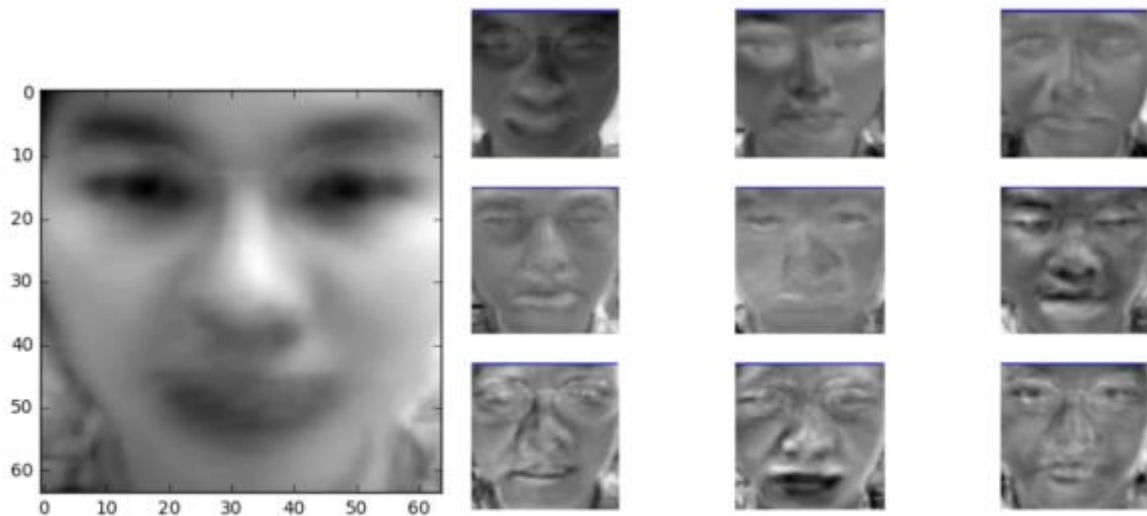


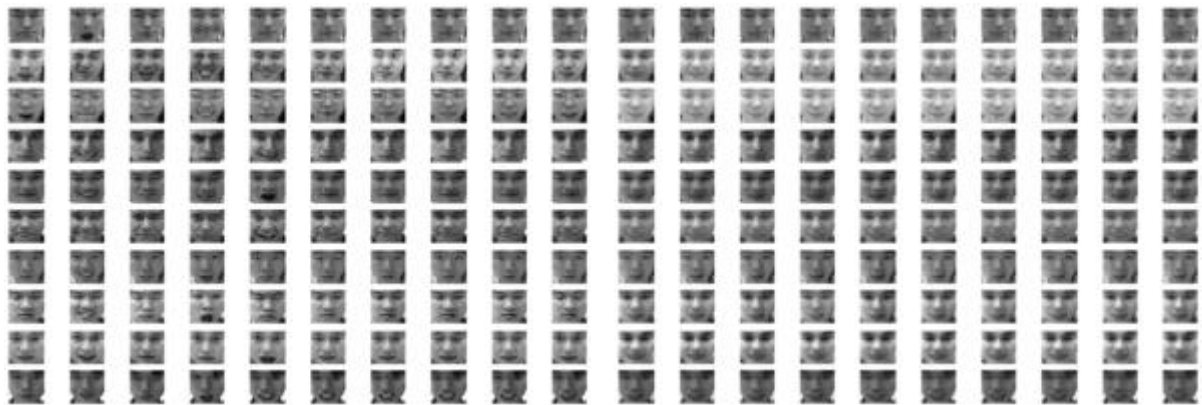
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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)

若除以 256 需要 59 張 eigenfaces

若除以 255 需要 60 張 eigenfaces

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：

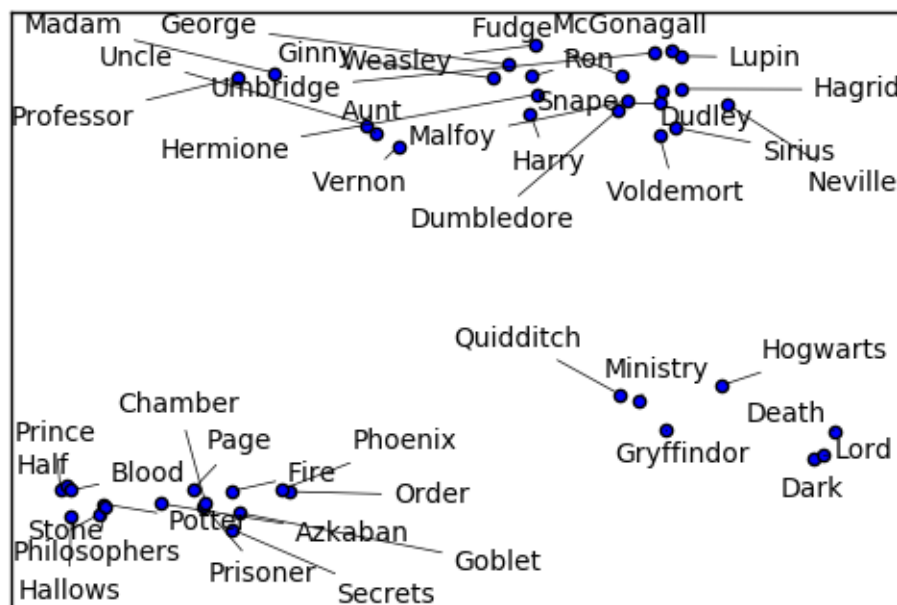
size: 每個詞向量的維度

window: 每個字之間最大會略過的長度

sample:每個詞出現的門檻，如果一個詞出現太多次的話會被取消取樣  
hs:有沒有用 hierarchical softmax，與 NPLM 相比可以大幅提升訓練速度  
negative:每次做負採樣取的單詞 negative example 的數量  
iter\_:迭代次數  
min\_count:出現次數小於這個數字的會被捨棄  
alpha:一開始的 learning rate  
cbow:是否使用 cbow 模型

## 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：

右上角的向量出現的大量的人名, 左上角則是人名的稱謂, 下方則是書中出現的各種特殊名詞。

## 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：

先產生一個資料集當作訓練 SVR 的資料，用這個 SVR 結果來猜 data.npz 的原始維度，找的 eigenvalues 裡，sample 代表要從這個資料集中用多少值來找附近的點，用這些點形成的平面來估計這筆資料集的原始維度應該是多少，這樣的方法只要預測的資料能過變成 array 就能夠帶入進行預測。

## 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：

預測原本的維度為 10，沒有答案無法確定結果正確與否，圖片本身為 480\*512，圖片能處理過(縮圖，擷取等等)再跑比較好，480\*512 的值過大有可能記憶體不足。