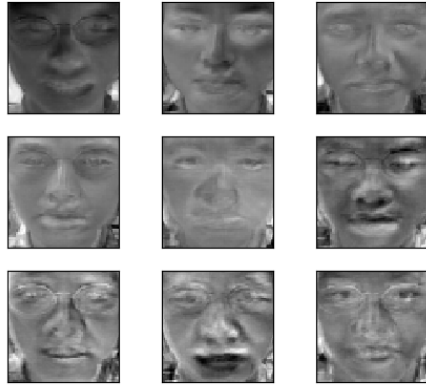


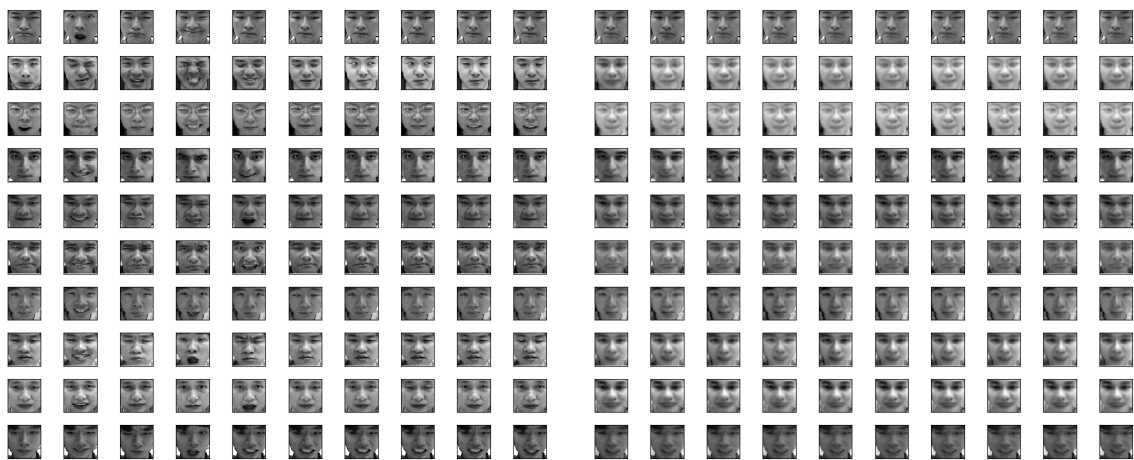
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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 = 59

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答：

size: word vector 的大小

window: 跳過的 word 長度

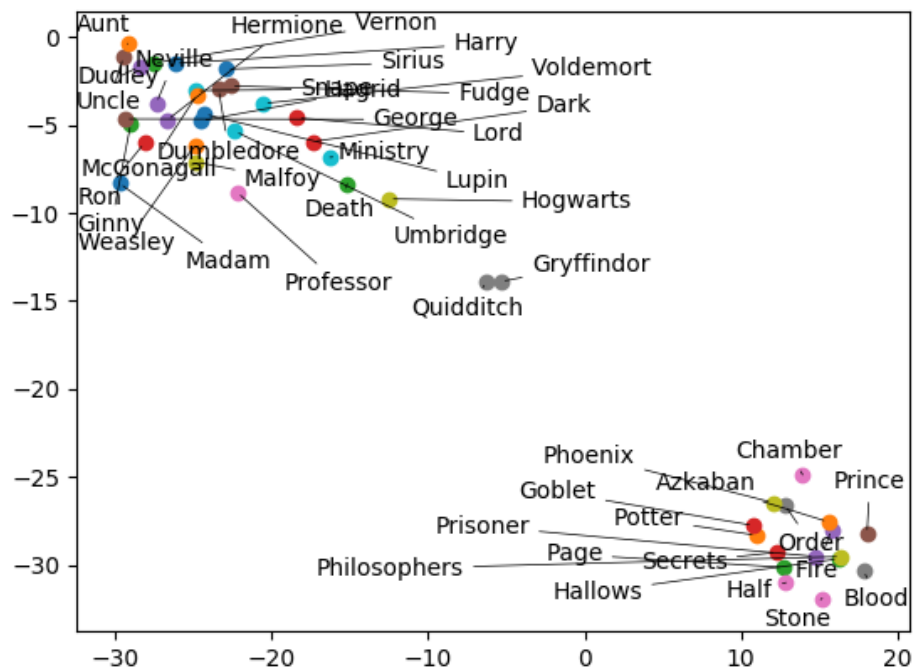
sample: 出現高於一個頻率的字彙，會被 down-sampled

min_count: 丟掉少於一個 count 的字彙。

alpha : Learning rate。

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: 幾個比較相關的字彙在附近出現, 像是 Ron 與他的妹妹 Ginny 還有她們的姓氏皆在附近。另外還有哈利的姨丈 Vernon 與他的表哥 Dudley 在附近, Uncle 和 Aunt 也很近原因可能是因為在文字中常常一起出現。

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:

運用原始維度 generate 出資料, 再從這些資料中算出 eigenvalue 值, 將這些 eigenvalue 當做要 training 的 data, 原始維度當做 label, 用 SVR (rbf kernel) model 下去 train, 把原本要求的 200 筆 set 當做 test Data, 從 train 完的 model 中去預測 dim。能得到答案是因為, eigenvalue 在 PCA 中有一定的特性, 可以當作 feature 來使用, 只要將任何資料算出 eigenvalue 丟進 model 中就能得到一個答案, 如果是從同樣的方法升維的話, 還會是對的。

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：

我的方法在這筆資料上，算不出 eigenvalue，因為這筆 dataset 的維度會有 (481, 245760)，但是理論上，我們無法得知這筆 data 是如何產生的，所以用原本的結果降低維度，所得到的答案，其實沒有意義。