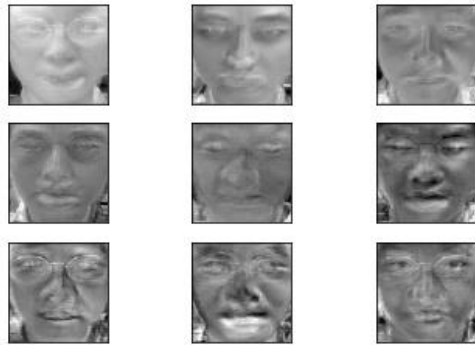


1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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：



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.

以 255 計算的話 $k = 60$

以 256 計算的話 $k = 59$

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

Size → 將一個詞轉換成多大的矩陣

Window → 決定要檢視前後多少個詞

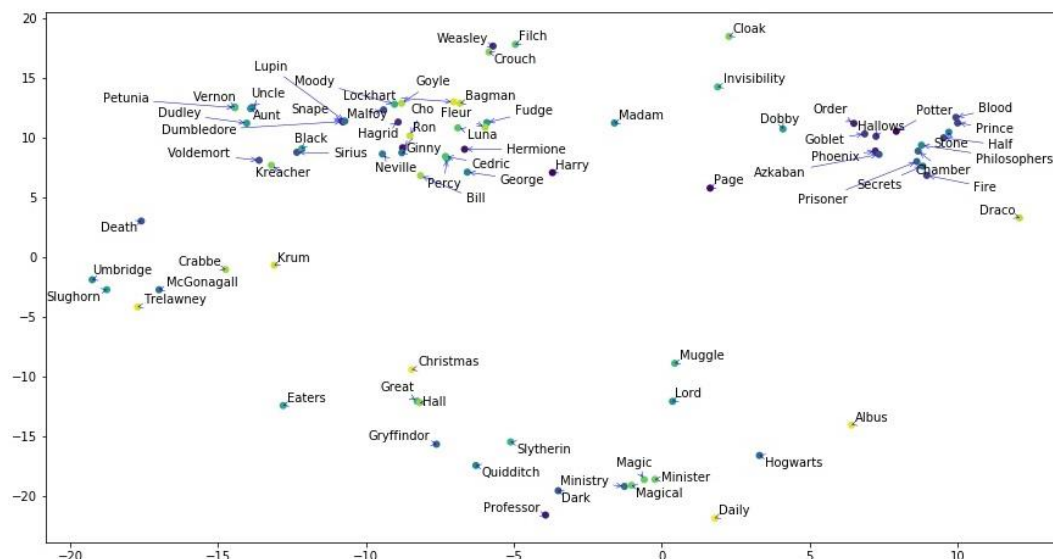
Alpha → 相當於 Learning Rate

Iter → Iterate 多少次

Min_count → 忽略詞頻少於此數的詞

Window 與 Size 大抵會決定訓練所花費的時間，過大的 Window 可能會產生過度解釋，過小則無法精準判斷詞意。

2.2. 將 word2vec 的結果投影到 2 維的圖:



2.3. 從上題視覺化的圖中觀察到了什麼？

答：

我發現此圖大致上能分成兩大部分，上方偏左的部分大多屬於書中主要腳色知人名，包括榮恩、妙麗、哈利等等。上方偏右之部分則很明顯的是代表每一集之書名。中間偏左一小區屬於霍格華茲高階的教授，例如校長、格來分多院長及史萊哲林院長等等。下方比較分散的一區則是與整個魔法世界和霍格華茲有關的設定，例如魔法部門、**Daily Prophet** 報紙、黑魔法、葛來芬多、魁地奇等等。

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：

首先我用教授提供之 **data generation** 生出 120000 個 **data set**，每一維度都有 2000 個 **data set**，而每一個 **set** 則有 4000 筆資料。我將此 4000 筆資料之平均以及變異數算出來，各成為一個 1*100 的向量。之後則用這兩個向量當成 **Input** (把兩個向量接再一起變成 1*200) 丟進 **neural network** 訓練，訓練的方式採用迴歸而非分類。我認為這個方法的通用性極差，由於 **training data** 是用特定的函式產生出來的，若一筆資料不是用此函式轉換維度，那我們的預測將很不準確。

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：