學號:B04705014 系級: 資管二 姓名:梁淯程

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:

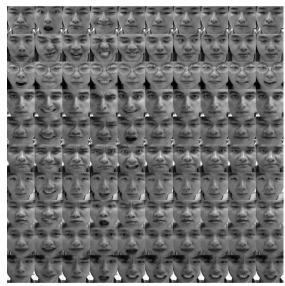
答:(左圖平均臉,右圖為 3x3 格狀 eigenfaces,順序為 左到右再上到下)





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 (測試結果如下圖)

```
54

(54, 4096)

0.0117688857812

55

(55, 4096)

0.0113896462274

56

(56, 4096)

0.0110199748664

57

(57, 4096)

0.0106537029158

58

(58, 4096)

0.01029988621

59

(59, 4096)

0.00996823197098

60

(60, 4096)

0.00967823027391

61

(61, 4096)

0.009938072504409

62
```

2.1. 使用 word2vec toolkit 的各個參數的值與其意義:

答: 行末的括弧裡是我使用的參數

size: word vector 的大小 (100)

window: 考慮一個詞的前 n 個和後 n 個 (15)

iter: 就是 epoch (20)

min_count: 出現低於 n 次的詞會被略過 (5)

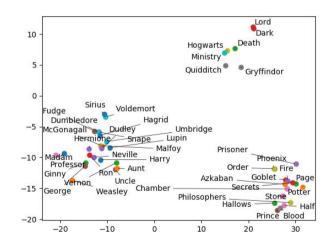
alpha: 就是 learning rate (0.01)

sample: 調高可以限制某個詞的出現次數,對常出現的詞減少 sample(0)

cbow: 使用 cbow 這個 model (預設是 skip-gram) (0)

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麽?

答:

人名(左下)以及書名(右下)很容易就全部聚再一起,至於右上角感覺是想要從 dark lord 延伸出一些東西,但是結果貌似不太好(dark lord 跟 gryffindor 很靠近)。

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:

首先生產出原始維度是 $1^{\sim}60$ 的 60 類 data (處理後的維度是 100),然後針對每一類,把處理後的 data 降維到 $1^{\sim}60$,再 reconstruct 回 100,計算 reconstruction loss,得到 60 組機率分佈 (組共 60*60 個數字),分別代表原始維度是 $n(n=1^{\sim}60)$ 降維到 $1^{\sim}60$ 再 reconstruct 回 100 的機率分佈 (通常離原始維度越近,loss 就會越小),接著再對 test data 計算新的機率分佈,看看跟哪個分佈比較接近,就猜那個維度。我認為這個方法對於依照助教的 code 生產出來的 test data,是非常合理的,畢竟 data 的性質都很相近,適用同樣的結論 (loss) 的機率分佈)。對於其他類型的 data,應該就沒辦法通用了。

3.2. 將你的方法做在 hand rotation sequence datatset 上得到什麼結果?合理嗎?請討論之。

答:

我認為我用上一題的方法估計出來的維度很有可能不太準確,原因是因為上一題我用的方法是直接計算 train data 的 reconstruction loss 分佈,理論上需要具有類似性質的 data 才能正常估計。