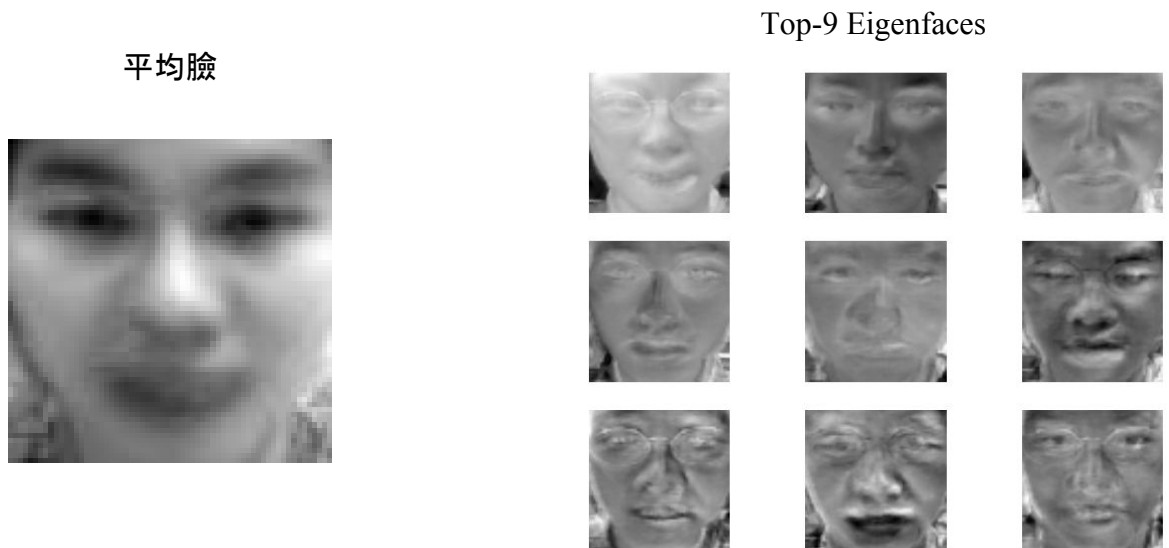


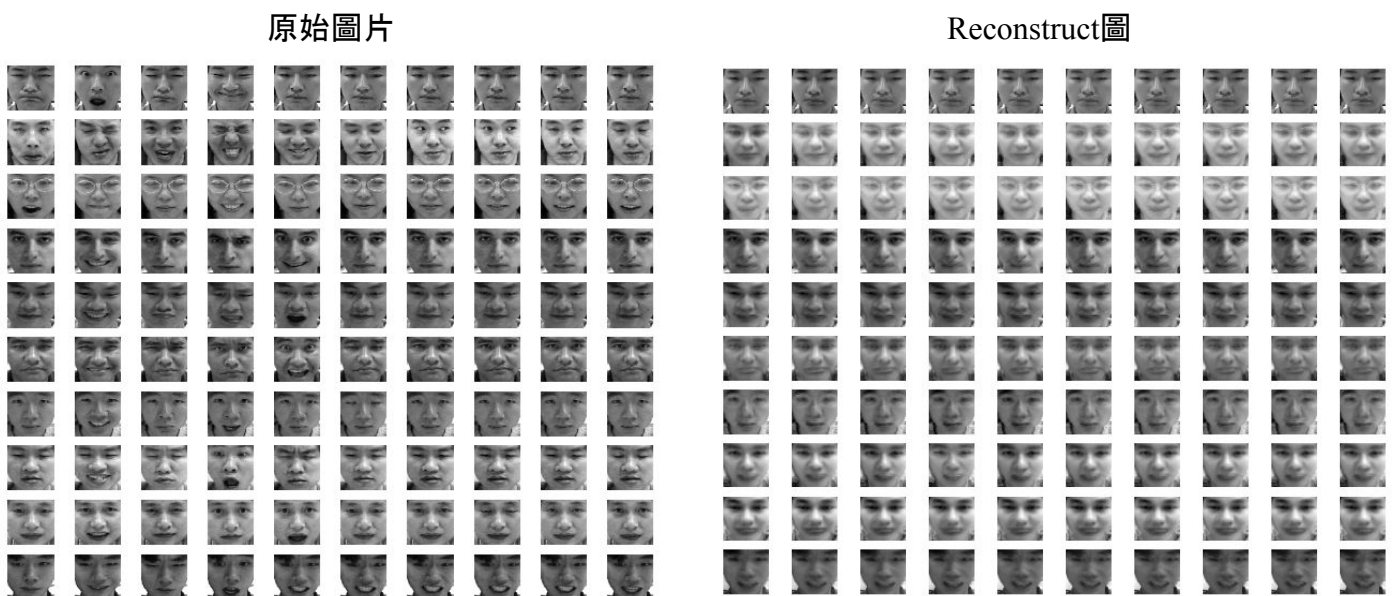
1.1. Dataset 中前 10 個人的前 10 張照片的平均臉和 PCA 得到的前 9 個 eigenfaces:

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

答：

**參數說明：**

negative = 隨機取樣Negative words的數目（通常在5~20之間）。

答：(使用TSNE將前1000多的單字投影到2維)



答：

有趣的是，Dursleys和Death Eaters非常的接近，也許表示在文中Dursleys和Death Eaters都代表著主角不幸的感覺。

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：

我的方法是利用轉換資料的函式皆為連續且可微的性質，假設每一個高維度的資料點和鄰近的資料點可以投影到同一個space(tangent space)。再加上透過觀察發現：如果從d維原始維度sample出來的資料點map到100維上，所取得的eigenvalues會在第d個和第d+1個有明顯的改變，而且每一個原始維度產生出來的eigenvalues分佈皆不太一樣。

因此我的作法如下：

1. 隨機取樣1/10的樣本點，計算每個樣本點和周圍的20-nearest neighbors。
2. 對每個樣本點及其鄰近點作PCA的投影。
3. 將eigenvalues進行normalization並由大排到小。
4. 平均各樣本點所得到的eigenvalues，得到100個平均eigenvalues。
5. 利用gen.py產生的資料點訓練svr模型(使用參數C=1.5)。  
(輸入為資料點的平均eigenvalues，預測目標為ln原始維度。)
6. 用svr來預測每筆測資的平均eigenvalues的分佈。

此方法的在kaggle上public的分數為0.。

此方法是透過eigenvalues的分佈作為代表資料點的原始維度，以此來估計原始維度。通用性上的限制是假設產生模擬資料的方式是從 $N(0, 1)$ 的分佈取樣出來的，若資料產生的方式和假設不符，就得使用其他方法來估計原始維度。

(例如：計算  $\frac{\sum_{i=1}^d \lambda_i}{\sum_{i=1}^{100} \lambda_i}$ ， ( $\lambda_i$  為ith eigenvalue)，找最小的d使之大於門檻值。)

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：