

# Assignment 2：人臉辨識

系級：資訊工程系碩士二年級

學號：610921213

姓名：陳怡蓁

使用環境：jupyter

## ● How you do the assignment

將 att\_faces 資料夾中的所有 pgm 影像檔讀取進來，並將每人取 5 張的圖片轉換成 1034\*1 的特徵向量，每個特徵向量即為一個樣本。分別為「是人臉」、「不是人臉」的樣本集。接下來我選擇使用 Bayes classifier 去做分類，利用 PCA 計算 200 張的圖像的轉換矩陣，將維度降為 10、20、30、40、50 維，這些降維後的樣本，利用 LDA 找出轉換矩陣，上面 2 種轉換矩陣使用 Bayes classifier 分類後，再分別統計出混淆矩陣跟正確率。

◇ 混淆矩陣(實際、預測)：

	實際 yes	實際 no
預測 yes	TP(true positive)	FP(false positive)
預測 no	FN(false negative)	TN(true negative)

◇ 正確率：

Accuracy = (TP+TN)/全部資料總數

## ● Result

PCA 的正確率：

訓練集正確率 = 0.9809523809523809	} 10 維
測試集正確率 = 0.9666666666666667	
訓練集正確率 = 0.9761904761904762	} 20 維
測試集正確率 = 0.9388888888888889	
訓練集正確率 = 0.9452380952380952	} 30 維
測試集正確率 = 0.9277777777777778	
訓練集正確率 = 0.9428571428571428	} 40 維
測試集正確率 = 0.9333333333333333	
訓練集正確率 = 0.9452380952380952	} 50 維
測試集正確率 = 0.9277777777777778	

LDA 正確率：

訓練集正確率 = 0.9976190476190476  
測試集正確率 = 0.9

PCA 和 LDA 的混淆矩陣：

```
PCA 混淆矩陣
----- Dimeniton : 1 -----
[[ 54  6]
 [ 0 120]]
----- Dimeniton : 2 -----
[[ 49 11]
 [ 0 120]]
----- Dimeniton : 3 -----
[[ 47 13]
 [ 0 120]]
----- Dimeniton : 4 -----
[[ 48 12]
 [ 0 120]]
----- Dimeniton : 5 -----
[[ 47 13]
 [ 0 120]]
LDA 混淆矩陣
----- Dimeniton : X -----
[[ 42 18]
 [ 0 120]]
```

- **Discussions on the results**

雖然使用 PCA 算法，但在降維的過程中有可能把具有明顯特徵的向量給模糊化，因此有可能丟失判斷重要的依據。而 LDA 有可能會過度擬和數據，然後也因為 LDA 最多只能降到  $k-1$  的維數，所以才先是用 PCA 降維再來使用 LDA。目前想到的方法為將維度不要降那麼多。

## ● Summary

這次的作業讓我更清楚的知道 PCA 和 LDA 的差別，2 者相同的地方在於都可以對數據進行降維，也符合高斯分布模型，而他們不同的點在於，LDA 是有監督的降維方式(數據帶有標籤)，除了只能降  $k-1$  個維度還可以用於分類，而 PCA 是無監督的降維方式(數據不帶標籤)也無限制降維。