

利用 whos 看 load 進的資料型態

```
>> whos
```

Name	Size	Bytes	Class	Attributes
PCA_Test_data	163x130	169520	double	
PCA_Train_data	163x130	169520	double	
test_label	130x1	6856	string	
train_label	130x1	6856	string	

AR face 的 dataset 中 train 與 test 的資料都有 10 個人，每個人有 13 張圖，一共 130 張圖。並且每一張圖有 163 個元素，並把它拉成一個維度。

在 Cluster k=2，並將資料讀進去後，一開始跑時發現出現很多 warning，並且準確率也很低。我也就一直卡在這個地方，之後經過學長的幫忙得知這是因為在計算中，Covariance matrix 的有些項會在迭代的過程中趨近 0，導致 $\det(\text{Covariance matrix})$ 變成 0，代表 Covariance matrix 沒有 inverse，讓接下來的計算變成 NaN，導致 model 無法繼續下去。

所以之後就依照助教提供的 sample 檔來進行修改並且實驗不同 loop 數得結果助教提供的 sample 解決 Covariance matrix 的有些項會在迭代的過程中趨近 0 的問題：

1. 在 gaussianND 的 function 中，在 Covariance matrix 上加一些小雜訊，加在 Covariance matrix 的對角線上，雜訊大小是根據相對於 Covariance matrix 的數據而定的，因為這邊的 Covariance matrix 值一開始很大，所以設為 3000000

```
% Calculate the multivariate gaussian.  
Sigma=Sigma+(eye(10)*3000000);  
pdf = 1 / sqrt((2*pi)^n * det(Sigma)) * exp(-1/2 * sum((meanDiff * inv(Sigma) .* meanDiff), 2));
```

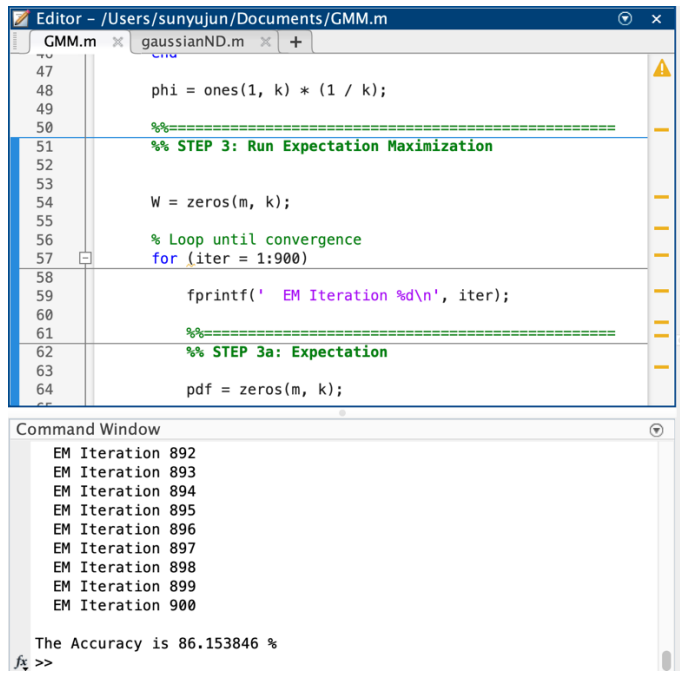
然後經過學長的幫助得知：

2. 在開始更新 mu 與 sigma 之前，在 X 加入一些小雜訊，也可以讓 variance 不會趨近 0，可以用 rand 的方式產生小雜訊。

```
% Subtract the cluster mean from all data points.  
% 加入雜訊 rand  
Xm = bsxfun(@minus, X+rand(13,10), mu(j, :));
```

接著就開始實驗不同的 loop 數與產生的準確率：

我測試到最好的準確率是在 loop 數控制為 900、Cluster k=2 時，準確率可以達到 86.153846%



The image shows a MATLAB Editor window with a script named GMM.m. The script includes comments and code for an Expectation-Maximization (EM) algorithm. The Command Window shows the output of the script, including the number of EM iterations and the final accuracy.

```
47
48     phi = ones(1, k) * (1 / k);
49
50     %%=====
51     %% STEP 3: Run Expectation Maximization
52
53
54     W = zeros(m, k);
55
56     % Loop until convergence
57     for (iter = 1:900)
58
59         fprintf(' EM Iteration %d\n', iter);
60
61         %%=====
62         %% STEP 3a: Expectation
63
64         pdf = zeros(m, k);
```

Command Window

```
EM Iteration 892
EM Iteration 893
EM Iteration 894
EM Iteration 895
EM Iteration 896
EM Iteration 897
EM Iteration 898
EM Iteration 899
EM Iteration 900

The Accuracy is 86.153846 %
fx >>
```

但是我測試在相同的 loop 數、Cluster 數時，跑的每一次都會產生不同的準確率，最後我想應該是在加入雜訊的時後是 rand 的，所以才會產生不同的準確率。