**Statistical Computing and Simulation** 統碩一 106354003 林健宏  
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

**Expectation-maximization algorithm**

(EM-Algorithm)

常用於機率模型中，尋找（1）參數使得likelihood function最大或是（2）最大後驗估計，而這個方法仰賴於無法觀測的隱性變量。

EM-Algorithm包含兩個步驟 : E – step與M – step

E- step

在給定當期參數下，計算整筆資料(包含隱性變量)的likelihood function對隱性變量的期望值。

M- step

讓E- step的期望值最大化，並令為下一個E- step的參數。

重複上述兩個動作，直到參數收斂。

本次使用資料為Mixture normal model的資料 – faithful，其中包含兩個觀測量：

eruptions: 噴發時間(以分為單位)；

waiting: 前一次噴發結束到下一次噴發開始的間隔時間(以分為單位)。

總共為272筆資料，下表僅呈現前5筆資料。

|  |  |
| --- | --- |
| eruption time(m) | waiting time(m) |
| 3.600 | 79 |
| 1.800 | 54 |
| 3.333 | 74 |
| 2.283 | 62 |
| 4.533 | 85 |

假定資料是由兩個二元常態的線性組合所構成(為eruption time，為waiting time)

其中除了 為觀測已知資料，其餘二元常態參數皆為未知，且。

隱性變量為, j = 1, 2, I = 1~272

使用EM-algorithm方法去找尋兩組二元常態參數

我們使用老師在講義推導的最後結果

定義



透過給定起始

經過R的運算後，結果如下

|  |  |  |  |
| --- | --- | --- | --- |
| t |  |  |  |
| 0 | (0.5,0.5) | (4, 60) |  |
| 1 | (0.3558696,0.6441304) | (2.036380, 54.478436) |  |
| 2 | (0.3558721,0.6441279) | (2.036387 ,54.478497) |  |
| 3 | (0.3558727,0.6441273) | (2.036388 ,54.478512) |  |
| 4 | (0.3558728,0.6441272) | (2.036388, 54.478515) |  |
| 5 | (0.3558728,0.6441272) | (2.036388, 54.478516) |  |
| 6 | (0.3558729,0.6441271) | (2.036388, 54.478516) |  |
| 7 | (0.3558729,0.6441271) | (2.036388, 54.478516) |  |
| 8 | (0.3558729,0.6441271) | (2.036388, 54.478516) |  |
| 9 | (0.3558729,0.6441271) | (2.036388, 54.478516) |  |
| 10 | (0.3558729,0.6441271) | (2.036388, 54.478516) |  |

(table 1)

|  |  |  |  |
| --- | --- | --- | --- |
| t |  |  |  |
| 0 | (0.5,0.5) | (3, 70) |  |
| 1 | (0.3558696,0.6441304) | (4.289655, 79.968030) |  |
| 2 | (0.3558721,0.6441279) | (4.289660 ,79.968090) |  |
| 3 | (0.3558727,0.6441273) | (4.289662 ,79.968110) |  |
| 4 | (0.3558728,0.6441272) | (4.289662, 79.968114) |  |
| 5 | (0.3558728,0.6441272) | (4.289662, 79.968115) |  |
| 6 | (0.3558729,0.6441271) | (4.289662, 79.968115) |  |
| 7 | (0.3558729,0.6441271) | (4.289662, 79.968115) |  |
| 8 | (0.3558729,0.6441271) | (4.289662, 79.968115) |  |
| 9 | (0.3558729,0.6441271) | (4.289662, 79.968115) |  |
| 10 | (0.3558729,0.6441271) | (4.289662, 79.968115) |  |

(table 2)

比較網路上現成的程式所得結果

（資料來源帶缺）

皆為迭代10次，起始參數皆設為一致

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 自己撰寫結果 | (0.3558729,0.6441271) | (2.036388, 54.478516) |  | (4.289662, 79.968115) |  |
| 網路運算結果 | (0.3558854,0.6441146) | (2.036421, 54.478880) |  | (4.289688, 79.968413) |  |

(table 3)

結論：從table 1以及table 2中看出，當t = 3之後各個參數都呈現收斂的狀態。我亦比較網路上他人所撰寫的程式結果，得到的結果相差無幾。

因此我推估這兩個二元常態為

, 其線性係數為0.3558729

, 其線性係數為0.6441271

Joint pdf:

Complete-data likelihood function:

=

令

E- step:

Q(θ|θ(t)) ≡ Ey [logp(x, y|θ)|x, θ(t)]