

繳交程式碼以及程式執行畫面

作業12

Ch04, Discrete HMM: Problem 1-Forward/Backward Algorithms

參考第四章講義32~40頁的
Forward與Backward Algorithms

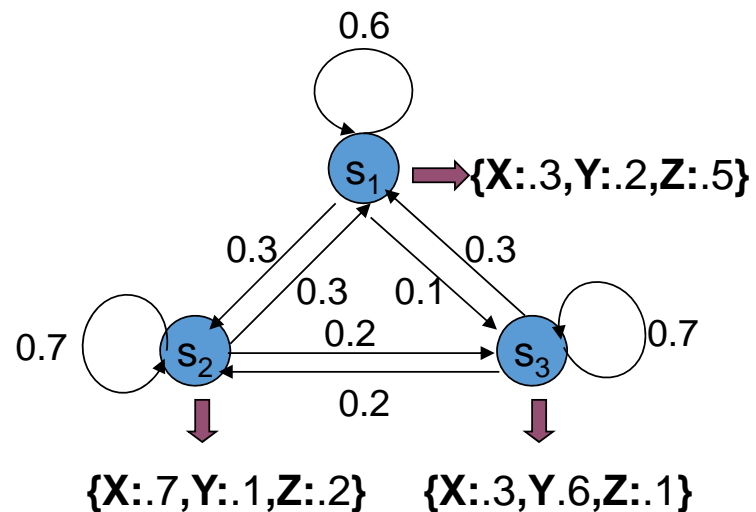
作業敘述

- 利用給予的三組HMM model以及三組observation，分別使用 **Forward Algorithm**和**Backward Algorithm** 來判斷三組observation分別在哪個Model的機率最大(程式執行圖如下)。

```
obser1
model_1 forward:4.230512e-19 backward:4.230512e-19
model_2 forward:1.334255e-29 backward:1.334255e-29
model_3 forward:5.132651e-25 backward:5.132651e-25
obser2
model_1 forward:1.438477e-29 backward:1.438477e-29
model_2 forward:1.635501e-17 backward:1.635501e-17
model_3 forward:2.515536e-05 backward:2.515536e-05
obser3
model_1 forward:2.700500e-20 backward:2.700500e-20
model_2 forward:2.700500e-20 backward:2.700500e-20
model_3 forward:1.861021e-19 backward:1.861021e-19
```

作業敘述

- 本作業HMM模型有三個state，每個state會有X、Y、Z三種可能的observation。



作業敘述

- HMM Model λ_1

A			B			π
0.2	0.7	0.1	0.5	0.4	0.1	0.7
0.1	0.2	0.7	0.7	0.2	0.1	0.2
0.7	0.1	0.2	0.7	0.1	0.2	0.1

作業敘述

- HMM Model λ_2

A			B			π
0.7	0.2	0.1	0.1	0.8	0.1	0.1
0.3	0.6	0.1	0.2	0.7	0.1	0.7
0.1	0.2	0.7	0.4	0.5	0.1	0.2

作業敘述

- HMM Model λ_3

A			B			π
0.2	0.7	0.1	0.1	0.2	0.7	0.2
0.6	0.3	0.1	0.2	0.2	0.6	0.2
0.2	0.7	0.1	0.3	0.1	0.6	0.6

作業敘述

- 三組observation分別為obser1.txt、 obser2.txt 、 obser3.txt，其中
obser1.txt是利用model1跑五十次所產生出來的觀察序列
(observation sequence)，相對的obser2.txt是利用model2、
obser3.txt是利用model3所跑出來的。
- 因此，obser i 在給定 λ_i 下，會有最大的機率（亦即 $P(\text{obser}i|\lambda_i) > P(\text{obser}i|\lambda_j), j \neq i$ ）， $i = 1, 2, 3$ 。

作業提示－模組使用

- 因為python對txt讀檔後會是string的data type所以需要進行格式的轉換
- 實驗室提供DSPbox 模組來讓同學使用str2ndar的功能(將string轉換成ndarray)
- Python import 原理: 當在Python import module時 import 陳述句會在相同目錄尋找 module.py，如果沒找到才會去sys.path中遞迴尋找 module.py，如果還是沒有，則會引發ImportError 例外。
- 在這裡我們建議同學將DSPbox.py放在與編寫程式相同的目錄來使用。

作業提示－模組使用

- `str2ndar`函式會將一組含有X、Y、Z的字串轉分別轉換成0、1、2，例如：字串內容為XYYZZ 經過`str2ndar`轉換後→ 01122，目的是方便同學對應到各model的observation矩陣。

- 模組使用範例：

```
>>> import DSPbox as dsp          #import DSPbox module as dsp
>>> f1 = open('obser1.txt','r')    # 先對obser1.txt進行開檔並設定為讀取模式
>>> obser1 = f1.read()             # 將obser1.txt進行讀取並存到obser1
>>> obser_new1 = dsp.str2ndar(obser1) # 將obser1進行格式轉換並
                                     #存到obser_new1裡
```

指令提示

- `print`指令結合變數輸出
- 使用方法`print ' %資料型態 ' %(對應的變數)`

範例:

```
>>> x =10
```

```
>>> print '5+5 = %.3e' %(x)
```

```
>>> 5+5=1.000e+01
```

- 作業中算出來的機率會使用科學符號表示，所以要使用`%e`
- 若使`%r`，則Python會自行選擇最適合的表示

議題討論

- Forward algorithm 是正確的機率值，而backward algorithm是在 independence assumption 的假設之下而有的演算法。
- 如果將forward和backward的機率輸出調到小數點後16位 (print '%.16e') 可發現forward和backward的機率其實是很些許不同的(如下圖)。

```
obser1  
model_1 forward:4.2305123481003905e-19 backward:4.2305123481003819e-19
```

- 從結果可知二者所求得的機率值可說是相同的，因此在這個例子中，independence assumption 的假設是成立的，也就是講義第38頁假設 B is independent of A 是成立的。