

需繳交

1.程式碼

2.訓練完的3個model (**model\_1.txt**, **model\_2.txt**, **model\_3.txt**)

3.在所得到的model下，使用Viterbi演算法對 Observations.txt 辨識的結果  
(**Observations\_Ans.txt**)

# 期中專案 — 離散HMM

Ch04, HMM: Problems 1,2,3

# 作業敘述 (1/4)

Problem 1所得到的 $\{\alpha_t(i), \beta_t(i)\}$ ，  
會在Problems 2,3使用到。

- 利用 training資料夾 內的三組training data (即model\_1\_training.txt, model\_2\_training.txt, model\_3\_training.txt)，分別訓練出3個HMM Model各自的 A, B,  $\pi$  (HMM Problem 3)。
- 檢驗模型是否訓練得準確，可使用Viterbi Algorithm對 check資料夾 內的testing data (包含測試資料test.txt與其對應的答案testAns.txt) 進行辨識 (HMM Problem 2)，test.txt的部分資料如下：

```
YZZZYXXZXZZZYZZYZXXXXZYZZZXZYZZXZXZXZXZZZXZZZZ
XZXYZZYXZZXZZXZZZZZZYZYZYZZZZYZZZYZZXZZZYZZYZX
YZZXZZXZYZXZXZYZZYZZZZZZZZXZYXZYZZXZYZZXZYZZYXZ
ZZYZZZYXZZZYZZYZXZZZZXZXZYZZXZZXZZZYZZYXZZZZX
XZYXXZXZXZZXZYZZYZXZYZZZXZZZZXZXZYZZYXXZYZZZZ
ZZZYXZZXZYZZYZZXZZZZYZZZZYZZXZZXZZXZZZYZZXZZXZZX
ZXZZXZYZZYXXZYZZYZZYXZZXZZZZZZXZYXZZZXZZXZZZZ
ZXZZZXZZYZZXZXZZYZZXZZZZZZXZYZZXZYZZZZXZZXZZ
```

....

→ 共有30列 (realizations,  $N_s = 100$ )，每一列是由 “  
某一個HMM模型” 所產生  
的一組sample序列  
(observation sequence)。

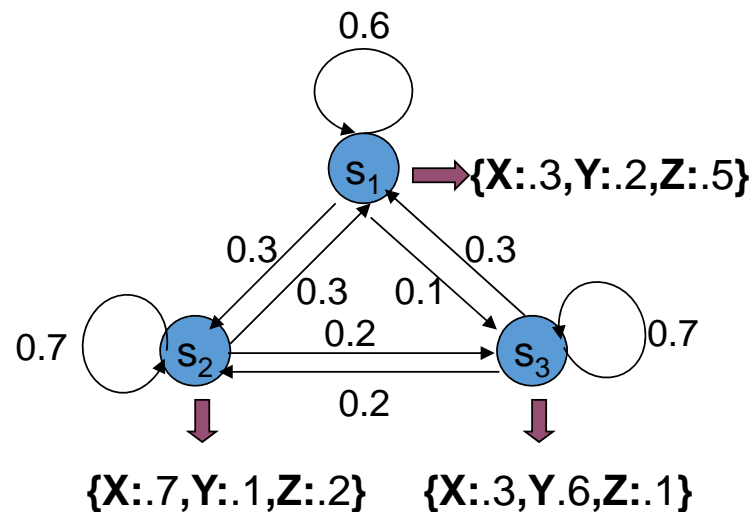
利用Viterbi Algorithm對  
testing data (即test.txt) 進  
行辨識後的正確答案在  
testAns.txt檔案內，並格  
式如下：

```
...
model 2
model 3
model 3
...
```

每一組sample序列 (observation sequence) 包含有50個observation ( $T = 50$ )。

## 作業敘述 (2/4)

- 本作業HMM模型都有三個state，每個state會有3種observation分別是X、Y、Z。



An example of 3 observation sequence

→ YXXXXXXXXXXXXYZZYXXXXXXXXXXXXZZYXXZXZXXZXXXYYYYXXZX

→ YXYXXYYXXYXXYXXYXXYXXZXXYXYYYYYYYYZYZYXXZYYYYYY

→ YXYYYYXYYYZZYZZYXXYYYYYZZYYYYYXYYXXXYXXYXXYZYY


## 作業敘述 (3/4)

- 所有HMM模型的初始化模型如下：

A				B			$\pi$	
	s1	s2	s3		x	y	z	
s1	0.3	0.3	0.4	s1	0.4	0.4	0.2	s1 0.3
s2	0.3	0.3	0.4	s2	0.3	0.3	0.4	s2 0.3
s3	0.4	0.4	0.2	s3	0.3	0.3	0.4	s3 0.4

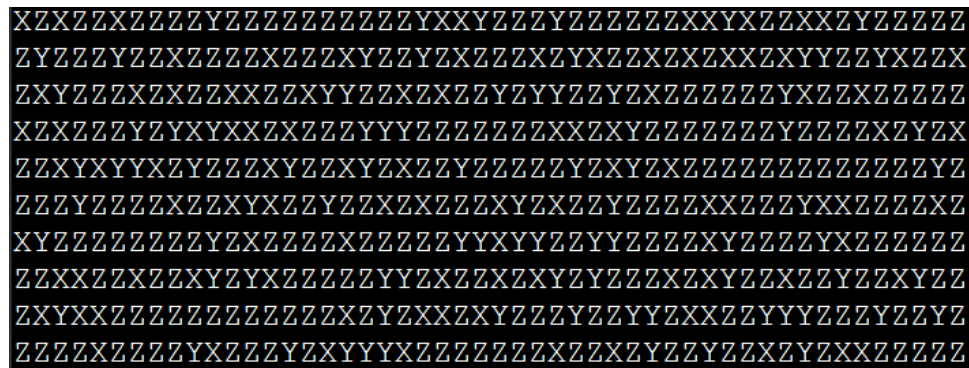
## 作業敘述 (4/4)

- 所給定之三組model的訓練資料 如下:



model\_1\_training.txt  
model\_2\_training.txt  
model\_3\_training.txt

- Training data (E.g., model\_1\_training.txt) 內容為100組Sample序列，一組有50個observation。



XZXZZXZZZZYZZZZZZZZZZZYXXYZZZYZZZZZZZXXYXZZXXZYZZZZZ  
ZYZZZYZZXZZZZXZZZXZYZZYZXZZXZZXZXZZXXZXYYZZZYXZZX  
ZXYZZZXZXZZXXZZXYYZZXZXZZYZYZZZYZXZZZZZZYXZZXZZZZZ  
XZXZZZYZYXYYXZXZZZYZZZZZZZZXXZXYZZZZZZZZYZZZZXZYZX  
ZZXYXYYXZYZZZXZYZZXZYZZYZZZZZYZXZYXZZZZZZZZZZZZYZ  
ZZZYZZZZXZZXYYZZYZZZXZXZZZXZYXZZYZZZZZXXZZZYXXZZZZXZ  
XYZZZZZZZZYZXZZZZXZZZZZYXYYZZYZZZZXZYZZZZYXZZZZZZ  
ZZXXZZXZZXYZYXZZZZZYZZXZZXZYZZZXZYZZXZZYZZZXYZZ  
ZXYXXZZZZZZZZZZZZXZYZZXZXZYZZYZZZYZZXXZZYZZYZZZYZ  
ZZZZXZZZZZYXZZZZYZXYYXZZZZZZZZXZZXZYZZYZZZXZYXZZZZZ

→ 共有100列 (realizations,  $N_s = 100$ )，每一列是由“某一個HMM模型”所產生的一組sample序列 (observation sequence)。

每一組sample序列 (observation sequence) 包含有50個observation ( $T = 50$ )。

# Training problem

$$\begin{aligned}\gamma_t(i) &= \sum_{j=1}^N \xi_t(i, j) \\ &= \frac{\xi_t(i, j)}{P(q_t = i, q_{t+1} = j, O | \lambda)} \\ &= \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O | \lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)}\end{aligned}$$

- 講義中的公式，是對一組 **Sample** 序列，如下：

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\bar{b}_j(k) = \frac{\sum_{\substack{t=1 \\ \mathbf{o}_t = \mathbf{v}_k}}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}$$

$$\bar{\pi}_i = \gamma_1(i)$$

# Training problem

- 而在作業會對所有 $N_s$ 組之Sample序列的訓練公式如下：

$$\bar{a}_{ij} = \frac{\sum_{n=1}^{N_s} \sum_{t=1}^{T-1} \xi_t^{(n)}(i,j)}{\sum_{n=1}^{N_s} \sum_{t=1}^{T-1} \gamma_t^{(n)}(i)} \quad \bar{b}_j = \frac{\sum_{n=1}^{N_s} \sum_{t=1}^T \gamma_t^{(n)}(j)}{\sum_{n=1}^{N_s} \sum_{t=1}^T \gamma_t^{(n)}(j)}$$
$$\bar{\pi}_i = \sum_{n=1}^{N_s} \gamma_1^{(n)}(i) / N_s$$

$N_s$ 是Sample序列 (realizations) 的數目

# 提示

- 在Training 過程中一次是不夠的，所以需設定iteration，其架構如下：

```
for iteration in range (0,10, 1 ):      # 10 次iteration
    for n in range (0,Ns,1):            # Ns lines (Ns sample序列)
        read n-th training
        compute  $\gamma^{(n)}$  ,  $\xi^{(n)}$ 
    end
    update A, B,  $\pi$ 
end
```



# 提示

- 在check資料夾中含有Testing Data (test.txt)，test.txt的答案在testAns.txt中，同學可依照辨識率來提高training iteration。

 test.txt	2017/5/8 下午 09...	文字文件	2 KB
 testAns.txt	2017/5/8 下午 09...	文字文件	1 KB

# Observations\_Ans.txt上傳格式

- 利用訓練出3個HMM Model，使用Viterbi演算法對 **Observations.txt** 裡的Sample進行辨識並輸出結果，輸出格式如下(檔名為 **Observations\_Ans.txt**)：

---

```
model 1  
model 1  
model 1  
model 1  
model 2  
model 2  
model 2  
model 2  
model 2  
model 2  
model 3  
model 3
```