

# INF281 Exercise 10 solutions

## 1. HMM probabilities

An HMM (hidden Markov model) is a probabilistic graphical model with three types of probabilities.

Transition probabilities:

	$L_t$	$H_t$
$L_{t-1}$	0.2	0.8
$H_{t-1}$	0.4	0.6

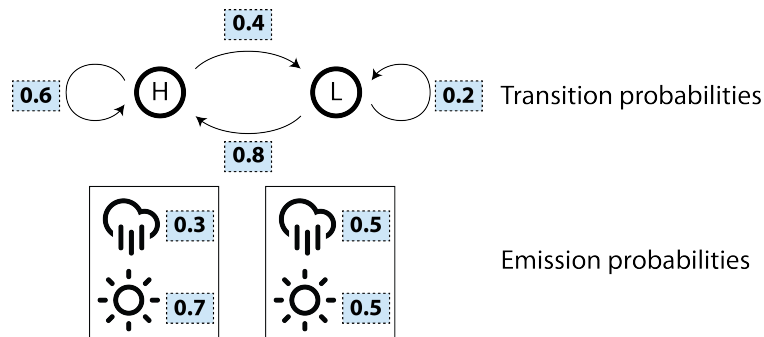
Emission probabilities:

	L	H
Sunny	0.5	0.7
Rain	0.5	0.3

Initial transition probabilities:

$$(L, H) = (0.3, 0.7)$$

- (a) Add the transition and emission probabilities to the graph.



- (b) What are the joint probabilities for (Rain, Rain, Sunny) and (H, L, L)?

**Solution:**  $p(H)p(\text{Rain}|H) \times p(L|H)p(\text{Rain}|L) \times p(L|L)p(\text{Sunny}|L)$   
 $= 0.7 \times 0.3 \times 0.4 \times 0.5 \times 0.2 \times 0.5 = 0.0042$

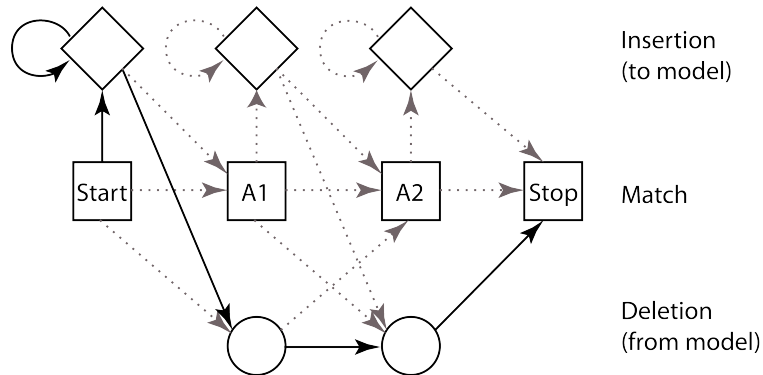
- (c) What are the joint probabilities for (Sunny, Rain, Sunny) and (L, H, L)?

**Solution:**  $p(L)p(\text{Sunny}|L) \times p(H|L)p(\text{Rain}|H) \times p(L|H)p(\text{Sunny}|L)$   
 $= 0.3 \times 0.5 \times 0.8 \times 0.3 \times 0.4 \times 0.5 = 0.0072$

## 2. HMM profile

A path of an HMM profile represents an alignment between an input sequence and the profile.

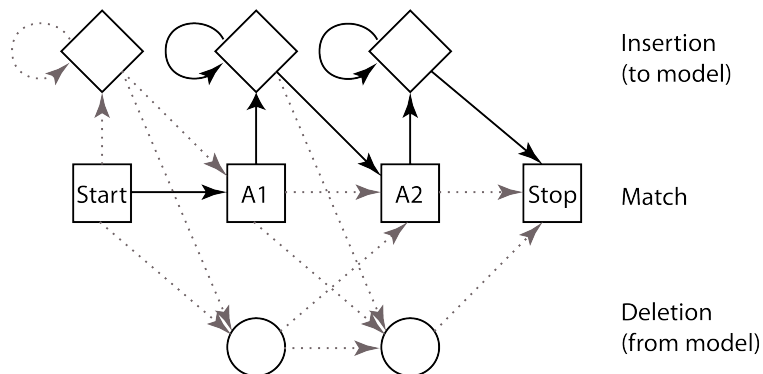
- (a) Assume Seq1 = q1 q2 and its path is indicated with solid lines. Draw the alignment of Seq1 and the profile.



### Solution:

Seq1: q1 q2 - -  
 Profile: - - A1 A2

- (b) Assume Seq2 = q1 q2 q3 q4 q5 q6 and its path is indicated with solid lines. Draw the alignment of Seq2 and the profile.

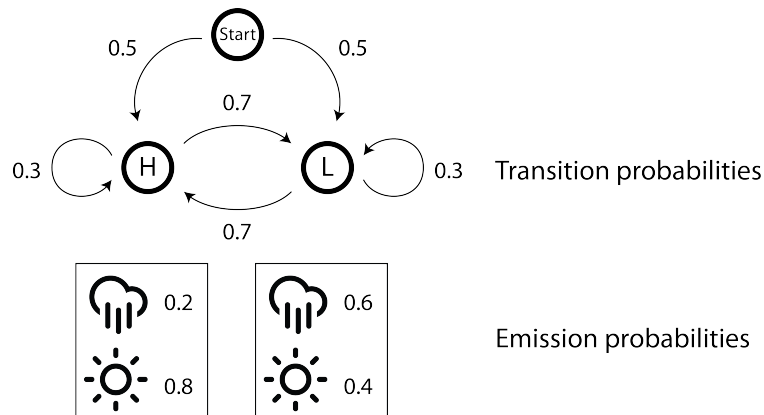


### Solution:

Seq2: q1 q2 q3 q4 q5 q6  
 Profile: A1 - - A2 - -

### 3. The Viterbi algorithm

The Viterbi algorithm is a dynamic programming based method to find the optimal path of an HMM with hidden status.



(a) Find the optimal path when observed weather conditions are (Rain, Sunny).

	H	L
Rain	$0.5 \times 0.2 = 0.1$	$0.5 \times 0.6 = 0.3$
Sunny	$(H) 0.1 \times 0.3 \times 0.8 = 0.024$ $(L) 0.3 \times 0.7 \times 0.8 = 0.168$	$(H) 0.1 \times 0.7 \times 0.4 = 0.028$ $(L) 0.3 \times 0.3 \times 0.4 = 0.036$

**Solution:** (L, H)

(b) Find the optimal path when observed weather conditions are (Sunny, Sunny, Rain).

	H	L
Sunny	$0.5 \times 0.8 = 0.4$	$0.5 \times 0.4 = 0.2$
Sunny	$(H) 0.4 \times 0.3 \times 0.8 = 0.096$ $(L) 0.2 \times 0.7 \times 0.8 = 0.112$	$(H) 0.4 \times 0.7 \times 0.4 = 0.112$ $(L) 0.2 \times 0.3 \times 0.4 = 0.024$
Rain	$(H) 0.112 \times 0.3 \times 0.2 = 0.007$ $(L) 0.112 \times 0.7 \times 0.2 = 0.016$	$(H) 0.112 \times 0.7 \times 0.6 = 0.047$ $(L) 0.112 \times 0.3 \times 0.6 = 0.02$

**Solution:** (L, H, L)

#### 4. The PROSITE language

The PROSITE language represents protein sequence patterns.

- x: An arbitrary amino acid
- -: Separating elements
- []: A list of amino acids
- {}: A list of not accepted amino acids
- (): A range of an element

Find all matched sequences for the following patterns. Assume the alphabet  $M = \{A, B, C\}$ .

(a)  $A - [BC] - \{BC\}$

**Solution:**

ABA, ACA

(b)  $A - B(1, 2)$

**Solution:**

AB, ABB

(c)  $A - x - C$

**Solution:**

AAC, ABC, ACC