

INF281 Exercise 10

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 |

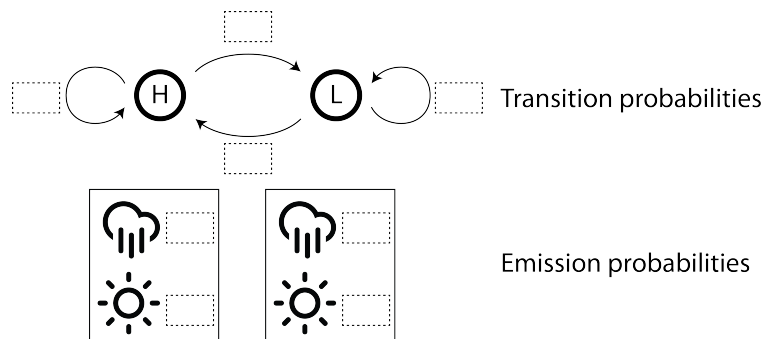
Emission probabilities:

| | 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.

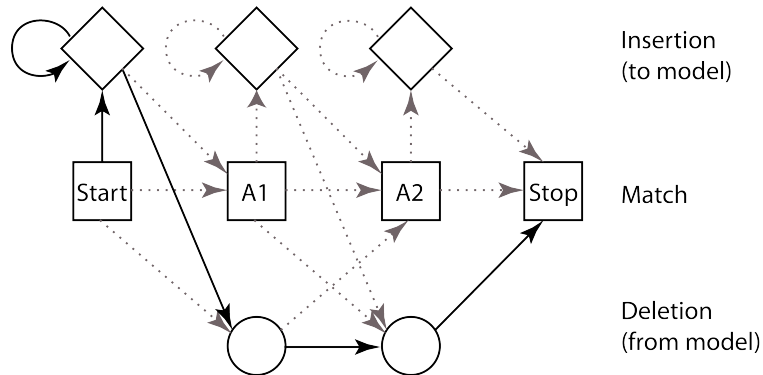


- (b) What are the joint probabilities for (Rain, Rain, Sunny) and (H, L, L)?
- (c) What are the joint probabilities for (Sunny, Rain, Sunny) and (L, H, L)?

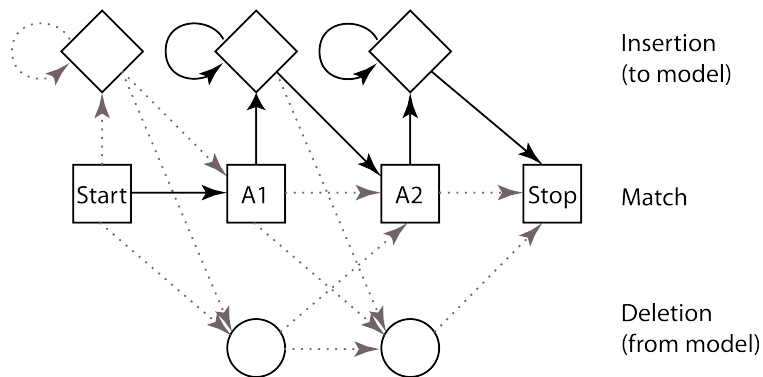
2. HMM profile

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

- (a) Assume $\text{Seq1} = q_1 q_2$ and its path is indicated with solid lines. Draw the alignment of Seq1 and the profile.

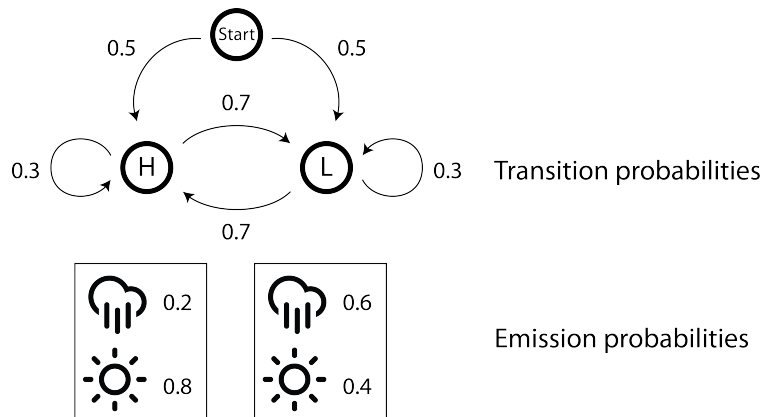


- (b) Assume $\text{Seq2} = q_1 q_2 q_3 q_4 q_5 q_6$ and its path is indicated with solid lines. Draw the alignment of Seq2 and the profile.



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 | | |
| Sunny | | |

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

| | H | L |
|-------|---|---|
| Sunny | | |
| Sunny | | |
| Rain | | |

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

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

(c) $A - x - C$