13 Hidden Markov model

13.1 Hidden Markov model

An HMM (hidden Markov model) is a probabilistic graphical model that assumes a Markov property. It contains two different types of probabilities: transition and emission probabilities.

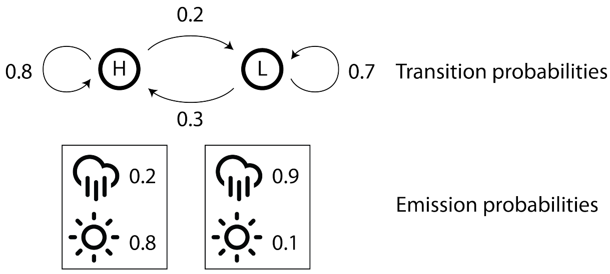

Figure 13.1: HMM for weather conditions

Example of HMM probability calculation

Calculate the probability when the observed weather conditions are (Sunny, Sunny, Sunny) and the corresponding states are (L, H, H). Assume no particular prior distribution for the initial states.

$$
\begin{aligned}
&p(\mathrm{L}) \cdot p(\mathrm{Sunny}|\mathrm{L}) \times p(\mathrm{H}|\mathrm{L}) \cdot p(\mathrm{Sunny}|\mathrm{H}) \times p(\mathrm{H}|\mathrm{H}) \cdot p(\mathrm{Sunny}|\mathrm{H}) \\
&= 0.5 \cdot 0.1 \times 0.3 \cdot 0.8 \times 0.8 \cdot 0.8 \\
&= 0.00768
\end{aligned}
$$

Search and training of HMM

A dynamic programming is commonly used to search the most probable path, and an EM (Expectation-Maximization) algorithm is often used for training.

- Viterbi algorithm: A dynamic programming for searching HMM
- BaumWelch algorithm: An EM algorithm for training HMM

Exercise 13.1

Use the transition and emission probabilities in the HMM above and calculate the probability when the observed weather conditions are (Rain, Rain, Sunny) and the corresponding states are (H, L, L).

13.2 Viterbi algorithm

The Viterbi algorithm is used to find the most probable path of HMM.

Probabilities of possible paths when the states are unknown

All possible paths need to be considered for an observed instance when the states are unknown.

Example of all possible paths

How many possible paths can one find when there are two states {S1, S2} and three observation {O1, O2, O3}?

The number of all possible paths: 8
(S1, S1, S1), (S1, S1, S2), (S1, S2, S1), (S1, S2, S2),
(S2, S1, S1), (S2, S1, S2), (S2, S2, S1), (S2, S2, S2)

Dynamic programming

The Viterbi algorithm is a dynamic programming that can be used to find the most probable path and its probability of an HMM.

Example of dynamic programming

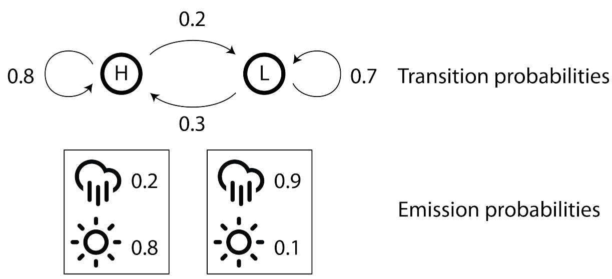

Figure 13.2: HMM for weather conditions

Find the most probable path in the HMM above when the observed weather conditions are (Sunny, Sunny, Sunny). Assume no particular prior distribution for the initial states.

Table 13.1: DP table for the Viterbi algorithm

	H	L
Sunny	$0.5 \times 0.8 = \mathbf{0.4}$	$0.5 \times 0.4 = 0.2$
Sunny	(H)$0.4 \times 0.8 \times 0.8 = \mathbf{0.256}$ (L)$0.2 \times 0.3 \times 0.8 = 0.048$	(H)$0.4 \times 0.2 \times 0.1 = 0.008$ (L)$0.2 \times 0.7 \times 0.1 = 0.014$
Sunny	(H)$0.256 \times 0.8 \times 0.8 = \mathbf{0.16384}$ (L)$0.014 \times 0.3 \times 0.8 = 0.00336$	(H)$0.256 \times 0.2 \times 0.1 = 0.00512$ (L)$0.014 \times 0.7 \times 0.1 = 0.00098$

Exercise 13.2

Use the HMM above and find the most probable path for the following weather conditions. Assume no particular prior distribution for the initial states.

1. (Sunny, Rain).

	H	L
Sunny		
Rain		

2. (Rain, Rain).

	H	L
Rain		
Rain		

13.3 HMM profile

An HMM (Hidden Markov model) profile is similar to a regular profile, but it is based on a probabilistic graphical model.

HMM profile to find sub-strings

An HMM profile represents position-specific probabilities of amino acids.

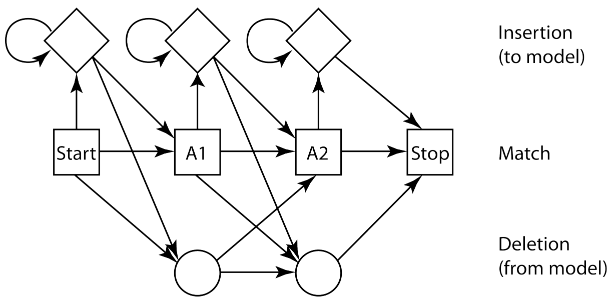

Figure 13.3: An HMM for an alignment of two columns A1 and A2

Example of HMM profile for finding sub-strings

Assume Seq1 = q1 q2 q3 q4 and its path is indicated with solid lines. Create the alignment of Seq1 and the profile.

	Start	Insertion	Match	Deletion	Stop
	-1				
q1		(2 start)			
q2			(4 deletion)	(3 insertion)	
q3		(5 match)			
q4		(6 Insertion)			
					(7 insertion)

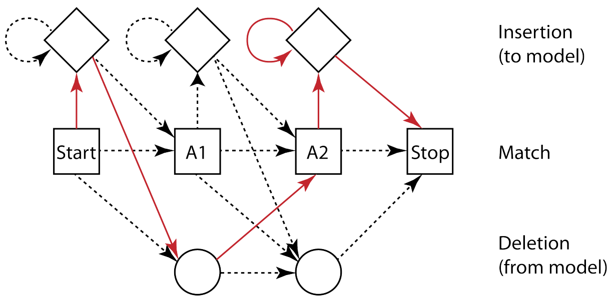

Figure 13.4: An HMM profile to find the optimal alignment

Local alignment:

```
q1  -   q2  q3  q4
-   A1  A2  -   -
```