CSE 411: Machine Learning

Hidden Markov Models

Dr Muhammad Abul Hasan

Department of Computer Science and Engineering
Green University of Bangladesh
muhammad.hasan@cse.green.edu.bd

Fall 2023

Outline

> *"If you are not willing to learn, no one can help you. If you are determined to learn, no one can stop you.*
>
> *– Zig Ziglar"*

Next Up ...

Probability Recap

- Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$
- Product rule: $P(x, y) = P(x|y)P(y)$
- Chain rule:
  $P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1) \ldots P(x_n|x_{n-1}, \ldots x_1)$
- x, y independent if and only if: $\forall x, y : P(x, y) = P(x)P(y)$

Reasoning over Time or Space

- Often, we want to reason about a sequence of observations
  - Speech recognition
  - Robot localization
  - User attention
  - Medical monitoring
- Need to introduce time (or space) into our models

Next Up ...

Markov Models

- Let us talk about the weather. Here in Bangladesh, we have three types of weather sunny, rainy, and foggy.
- Let us assume for the moment that the same weather lasts for the full day. i.e., it doesn't change from rainy to sunny in the middle of the day.

Weather prediction is all about trying to guess what the weather will be like tomorrow based on a history of observations of weather.

Markov Models

- Let us assume we would like to predict the weather for tomorrow.
- According to the chain rule of probability, tomorrow's weather depends on today, yesterday, until day one.
- We want to compute the following probability:

$$P(w_n|w_{n-1}, w_{n-2}, \ldots w_1). \tag{1}$$

Markov Models

- Using equation 1, we can compute the probabilities of types of weather for tomorrow and the next days using n days of history.
- For example, if we knew that the weather for the past three days was $\{\text{sunny}, \text{sunny}, \text{foggy}\}$ in chronological order the probability that tomorrow would be rainy is given by:

$$P(w_4 = Rainy|w_3 = Foggy, w_2 = Sunny, w_1 = Sunny) \tag{2}$$

Markov Model

- Here is the problem the larger n is the more statistics we must collect.
- Suppose that $n = 100$, then we must collect statistics for past history of 100 days and model it for computation.
- However, we have good news. We can simplify the computation using an assumption called the Markov Assumption.
  In a sequence $\{w_1, w_2, \dots w_n\}$, according to Markov assumption:

$$P(w_n | w_{n-1}, w_{n-2}, \dots w_1) \approx P(w_n | w_{n-1}) \tag{3}$$

Markov Model

- Equation (3) is called a first-order Markov assumption since we say that the probability of an observation at time n only depends on the observation at time $n-1$.
- A second order Markov assumption would have the observation at time n depending on $n-1$ and $n-2$.

$$P(w_n|w_{n-1}, w_{n-2}, \dots w_1) \approx P(w_n|w_{n-1}, w_{n-2}) \tag{4}$$

- In general, when people talk about Markov assumptions they usually refer to first-order Markov assumptions.

Markov Models

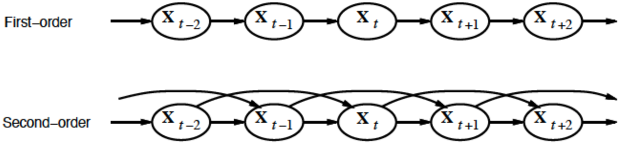
- We can express the joint probability using the Markov assumption:

$$P(w_1 \cdots w_n) = \prod_{i=1}^{n} P(w_i | w_{i-1}) \tag{5}$$

- So this now has a profound effect on the number of histories that we have to find statistics for – we now only need $3^2 = 9$ numbers to characterize the probabilities of all of the sequences.
- This assumption may or may not be a valid assumption depending on the situation (in the case of weather, it's probably not valid) but we use these to simplify the situation.

Markov processes (Markov chains)

- **Markov assumption**: X_t depends on *bounded* subset of $X_{0:t-1}$
- **First-order Markov process**: $p(X_t|X_{0:t-1}) = p(X_t|X_{t-1})$
- **Second-order Markov process**: $p(X_t|X_{0:t-1}) = p(X_t|X_{t-2}, X_{t-1})$



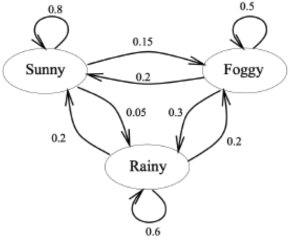
Initial Statistics

- As we are modeling Bangladeh's weather, we need historical data.
- Let us assume that the history of weather is as follows:
  - January 1, 2023 - Sunny
  - January 2, 2023 - Sunny
  - January 3, 2023 - Foggy
  - ...
  - May 31, 2023 - Rainy
  - ...
  - December 31, 2023 - Foggy

Weather is a continuous system that produces sequential data.

Markov Models

- From the sequential data given in the previous slide, we can generate a diagram as follows.



- The diagram helps computing probability expressed as $P(w_{tomorrow}|w_{today})$.
- For example $P(w_{tomorrow} = \text{rainy}|w_{today} = \text{sunny})$ is 0.05.
- Similarly $P(w_{tomorrow} = \text{rainy}|w_{today} = \text{rainy})$ is 0.6.

Markov Models

- For first-order Markov models, we can use these probabilities to draw a probabilistic finite state automaton.
- For the simplified weather domain, you have three states: {Sunny, Rainy, and Foggy} and every day we would transition to a (possibly) new state based on the probabilities given in the diagram shown in one of the previous slides.
- Let us fill up the following table based on the information given in that diagram.

Sunny	Sunny	Rainy	Foggy
Rainy			
Foggy			

- The above tabular information is called **Transition Probability Table**.

Next Up ...

Question 1

1. Given that today is **Sunny** what is the probability that tomorrow is **Sunny** and the day after is **Rainy**?

Answer 1

- We need to calculate $P(w_2 = S, w_3 = R | w_1 = S) =?$

$$
\begin{aligned}
& P(w_2 = S, w_3 = R | w_1 = S) \\
= \quad & \underbrace{P(w_3 = R | w_2 = S, w_1 = S)}_{\text{Apply Markov Assumption}} \times P(w_2 = S | w_1 = S) \\
= \quad & P(w_3 = R | w_2 = S) \times P(w_2 = S | w_1 = S) \\
= \quad & 0.05 \times 0.8 \\
= \quad & 0.04
\end{aligned}
$$

Answer 1

- We need to calculate $P(w_2 = S, w_3 = R | w_1 = S) = ?$

$$
\begin{aligned}
& P(w_2 = S, w_3 = R | w_1 = S) \\
= \; & \underbrace{P(w_3 = R | w_2 = S, w_1 = S)}_{\text{Apply Markov Assumption}} \times P(w_2 = S | w_1 = S) \\
= \; & P(w_3 = R | w_2 = S) \times P(w_2 = S | w_1 = S) \\
= \; & 0.05 \times 0.8 \\
= \; & 0.04
\end{aligned}
$$

Question 2

2. Given that today is **Sunny** what is the probability that tomorrow is **Foggy** and the day after tomorrow is **Rainy**?

Answer 2.

Please do it yourself.
Answer is: 0.045

Question 3

3. Given that today is **Sunny** what is the probability that the day after tomorrow is **Rainy**?

Answer 3

The answer is a bit tricky. As tomorrow's weather state is not mentioned, you need to consider all three states for that to compute the questioned probability.

$$
\begin{aligned}
& P(w_3 = R, w_2 = ?|w_1 = S) \\
= \quad & P(w_3 = R|w_2 = S, w_1 = S) \times P(w_2 = S|w_1 = S) \\
+ \quad & P(w_3 = R|w_2 = R, w_1 = S) \times P(w_2 = R|w_1 = S) \\
+ \quad & P(w_3 = R|w_2 = F, w_1 = S) \times P(w_2 = F|w_1 = S) \\
= \quad & P(w_3 = R|w_2 = S) \times P(w_2 = S|w_1 = S) \\
+ \quad & P(w_3 = R|w_2 = R) \times P(w_2 = R|w_1 = S) \\
+ \quad & P(w_3 = R|w_2 = F) \times P(w_2 = F|w_1 = S) \\
= \quad & 0.05 \times 0.8 + 0.6 \times 0.05 + 0.3 \times 0.15 \\
= \quad & 0.04 + 0.03 + 0.045 \\
= \quad & 0.115
\end{aligned}
$$

Question 4

4. Given that today is **Foggy** what is the probability that it will be **Rainy** two days from now?

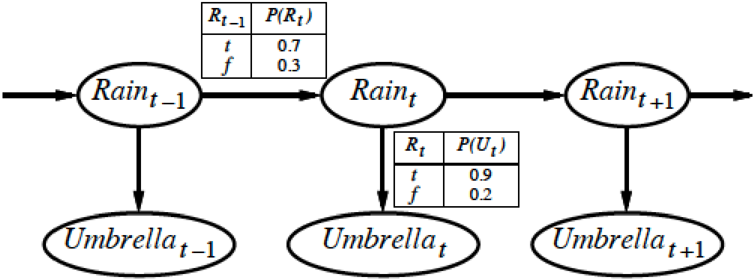
Answer 4

Please do it yourself.
Answer is: 0.34

Next Up ...

Hidden Markov Model (HMM)

HMM is a special type of Bayes net, X_t is single discrete random variable:



with joint probability distribution

$$\mathbf{P}(X_{0:t}, E_{1:t}) = \mathbf{P}(X_0) \prod_{i=1}^{t} \mathbf{P}(X_i | X_{i-1}) \mathbf{P}(X_i | E_i)$$

Hidden Markov Model: Example

- So what makes a Hidden Markov Model?
- Suppose you were locked in a room for several days and you were asked about the weather outside. The only piece of evidence that you have is whether the person who comes into the room carrying your daily meal is carrying an umbrella or not.
- Let us assume the following probabilities:

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Hidden Markov Model: Example

- So what makes a Hidden Markov Model?
- Suppose you were locked in a room for several days and you were asked about the weather outside. The only piece of evidence that you have is whether the person who comes into the room carrying your daily meal is carrying an umbrella or not.
- Let us assume the following probabilities:

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Hidden Markov Model: Example

- The equation for the weather Markov process, before you were locked in the room, was:

$$P(w_1, \ldots .w_n) = \prod_{i=1}^{n} P(w_i | w_{i-1}) \tag{6}$$

Hidden Markov Model: Example

- Now we have to factor in the fact that the actual weather is hidden from us.
- We do that by using Bayes' Rule:

$$P(w_1, \ldots, w_n | u_1, \ldots, u_n) = \frac{P(u_1, \ldots, u_n | w_1, \ldots, w_n) P(w_1, \ldots, w_n)}{P(u_1, \ldots, u_n)} \quad (7)$$

where

- u_i is $true$ if your caretaker brought an umbrella on day i and $false$ if the caretaker didnot.
- The probability $P(w_1, \ldots, w_n)$ is the same as the Markov model from the last section, and
- the probability $P(u_1, \ldots, u_n)$ is the prior probability of seeing a particular sequence of umbrella events (e.g., $\{True, False, True\}$).
- The probability $P(u_1, \cdots, u_n | w_1, \cdots, w_n)$ can be estimated as $\prod_{i=1}^{n} P(u_i | w_i)$ if you assume that for all i given w_i, u_i is independent of all u_j and w_j for all $j \neq i$, we call it **emission probability**.

Hidden Markov Model: Example

- Now we have to factor in the fact that the actual weather is hidden from us.
- We do that by using Bayes' Rule:

$$P(w_1, \ldots, w_n | u_1, \ldots, u_n) = \frac{P(u_1, \ldots, u_n | w_1, \ldots, w_n) P(w_1, \ldots, w_n)}{P(u_1, \ldots, u_n)} \quad (7)$$

where

- u_i is $true$ if your caretaker brought an umbrella on day i and $false$ if the caretaker didnot.
- The probability $P(w_1, \ldots, w_n)$ is the same as the Markov model from the last section, and
- the probability $P(u_1, \ldots, u_n)$ is the prior probability of seeing a particular sequence of umbrella events (e.g., $\{True, False, True\}$).
- The probability $P(u_1, \cdots, u_n | w_1, \cdots, w_n)$ can be estimated as $\prod_{i=1}^{n} P(u_i | w_i)$ if you assume that for all i given w_i, u_i is independent of all u_j and w_j for all $j \neq i$, we call it **emission probability**.

Hidden Markov Model: Example

- Now we have to factor in the fact that the actual weather is hidden from us.
- We do that by using Bayes' Rule:

$$P(w_1, \ldots, w_n | u_1, \ldots, u_n) = \frac{P(u_1, \ldots, u_n | w_1, \ldots, w_n) P(w_1, \ldots, w_n)}{P(u_1, \ldots, u_n)} \tag{7}$$

where

- u_i is $true$ if your caretaker brought an umbrella on day i and $false$ if the caretaker didnot.
- The probability $P(w_1, \ldots, w_n)$ is the same as the Markov model from the last section, and
- the probability $P(u_1, \ldots, u_n)$ is the prior probability of seeing a particular sequence of umbrella events (e.g., $\{True, False, True\}$).
- The probability $P(u_1, \cdots, u_n | w_1, \cdots, w_n)$ can be estimated as $\prod_{i=1}^{n} P(u_i | w_i)$ if you assume that for all i given w_i, u_i is independent of all u_j and w_j for all $j \neq i$, we call it **emission probability**.

Next Up ...

HMM in Toy Exercise

Let us use the theory in practice.

Question 5

- Suppose the day you were locked in it was **Sunny**. The next day the caretaker carried an umbrella into the room. Assuming that the prior probability of the caretaker carrying an umbrella on any day is $\frac{1}{2}$, what is the probability that the second day was **Rainy**?

Solution 5

- We need to calculate $P(w_2 = R | w_1 = S, u_2 = Tr) = ?$.

$$
\begin{aligned}
P(w_2 = R | w_1 = S, u_2 = Tr) &= \frac{P(w_2 = R, w_1 = S | u_2 = Tr)}{P(w_1 = S | u_2 = Tr)} \\
(u_2 \ and \ w_1 independent) &= \frac{P(w_2 = R, w_1 = S | u_2 = Tr)}{P(w_1 = S)} \\
(Bayes' Rule) &= \frac{P(u_2 = Tr | w_1 = S, w_2 = R) P(w_2 = R, w_1 = S)}{P(w_1 = S) P(u_2 = Tr)} \\
(\text{Markov assumption}) &= \frac{P(u_2 = Tr | w_2 = R) P(w_2 = R, w_1 = S)}{P(w_1 = S) P(u_2 = Tr)}
\end{aligned}
$$

Solution 5

- We need to calculate $P(w_2 = R|w_1 = S, u_2 = Tr) =?$.

$$
\begin{aligned}
P(w_2 = R|w_1 = S, u_2 = Tr) &= \frac{P(w_2 = R, w_1 = S|u_2 = Tr)}{P(w_1 = S|u_2 = Tr)} \\
(u_2 \ and \ w_1 independent) &= \frac{P(w_2 = R, w_1 = S|u_2 = Tr)}{P(w_1 = S)} \\
(Bayes' Rule) &= \frac{P(u_2 = Tr|w_1 = S, w_2 = R)P(w_2 = R, w_1 = S)}{P(w_1 = S)P(u_2 = Tr)} \\
(\text{Markov assumption}) &= \frac{P(u_2 = Tr|w_2 = R)P(w_2 = R, w_1 = S)}{P(w_1 = S)P(u_2 = Tr)}
\end{aligned}
$$

Solution 5

- We need to calculate $P(w_2 = R | w_1 = S, u_2 = Tr) = ?$.

$$\begin{aligned}
P(w_2 = R | w_1 = S, u_2 = Tr) &= \frac{P(w_2 = R, w_1 = S | u_2 = Tr)}{P(w_1 = S | u_2 = Tr)} \\
(u_2 \ and \ w_1 independent) &= \frac{P(w_2 = R, w_1 = S | u_2 = Tr)}{P(w_1 = S)} \\
(Bayes' Rule) &= \frac{P(u_2 = Tr | w_1 = S, w_2 = R) P(w_2 = R, w_1 = S)}{P(w_1 = S) P(u_2 = Tr)} \\
(\text{Markov assumption}) &= \frac{P(u_2 = Tr | w_2 = R) P(w_2 = R, w_1 = S)}{P(w_1 = S) P(u_2 = Tr)}
\end{aligned}$$

Solution 5

- We need to calculate $P(w_2 = R|w_1 = S, u_2 = Tr) =?$.

$$
\begin{aligned}
P(w_2 = R|w_1 = S, u_2 = Tr) &= \frac{P(w_2 = R, w_1 = S|u_2 = Tr)}{P(w_1 = S|u_2 = Tr)} \\
(u_2 \ and \ w_1 independent) &= \frac{P(w_2 = R, w_1 = S|u_2 = Tr)}{P(w_1 = S)} \\
(Bayes' Rule) &= \frac{P(u_2 = Tr|w_1 = S, w_2 = R)P(w_2 = R, w_1 = S)}{P(w_1 = S)P(u_2 = Tr)} \\
(\text{Markov assumption}) &= \frac{P(u_2 = Tr|w_2 = R)P(w_2 = R, w_1 = S)}{P(w_1 = S)P(u_2 = Tr)}
\end{aligned}
$$

Solution 5

- We need to calculate $P(w_2 = R|w_1 = S, u_2 = Tr) =?$.

$$\begin{aligned}
P(w_2 = R|w_1 = S, u_2 = Tr) &= \frac{P(w_2 = R, w_1 = S|u_2 = Tr)}{P(w_1 = S|u_2 = Tr)} \\
(u_2 \ and \ w_1 independent) &= \frac{P(w_2 = R, w_1 = S|u_2 = Tr)}{P(w_1 = S)} \\
(Bayes' Rule) &= \frac{P(u_2 = Tr|w_1 = S, w_2 = R)P(w_2 = R, w_1 = S)}{P(w_1 = S)P(u_2 = Tr)} \\
(\text{Markov assumption}) &= \frac{P(u_2 = Tr|w_2 = R)P(w_2 = R, w_1 = S)}{P(w_1 = S)P(u_2 = Tr)}
\end{aligned}$$

Solution 4 contd.

$$
\begin{aligned}
(P(A,B) = P(A|B)P(B)) &= \frac{P(u_2 = Tr|w_2 = R)P(w_2 = R|w_1 = S)P(w_1 = S)}{P(w_1 = S)P(u_2 = Tr)} \\
(Cancel : P(w_1 = S)) &= \frac{P(u_2 = Tr|w_2 = R)P(w_2 = R|w_1 = S)}{P(u_2 = Tr)} \\
&= \frac{0.8 \times 0.05}{0.5} \\
&= 0.08
\end{aligned}
$$

Solution 4 contd.

$$
\begin{aligned}
(P(A,B) = P(A|B)P(B)) &= \frac{P(u_2 = Tr|w_2 = R)P(w_2 = R|w_1 = S)P(w_1 = S)}{P(w_1 = S)P(u_2 = Tr)} \\
(Cancel : P(w_1 = S)) &= \frac{P(u_2 = Tr|w_2 = R)P(w_2 = R|w_1 = S)}{P(u_2 = Tr)} \\
&= \frac{0.8 \times 0.05}{0.5} \\
&= 0.08
\end{aligned}
$$

Question 6

- Suppose the day you were locked in the room it was **Sunny**; the caretaker brought in an umbrella on day 2, but not on day 3. Again assuming that the prior probability of the caretaker bringing an umbrella is $\frac{1}{2}$, what is the probability that it is **Foggy** on day 3?

Solution 6

- We need to calculate $P(w_3 = F|w_1 = S, u_2 = Tr, u_3 = Fl) = ?$

$$
\begin{aligned}
& P(w_3 = F|w_1 = S, u_2 = Tr, u_3 = Fl) \\
= \quad & P(w_2 = F, w_3 = F|w_1 = S, u_2 = Tr, u_3 = Fl) \\
& + P(w_2 = R, w_3 = F|w_1 = S, u_2 = Tr, u_3 = Fl) \\
& + P(w_2 = S, w_3 = F|w_1 = S, u_2 = Tr, u_3 = Fl) \\
= \quad & \frac{P(u_3 = Fl|w_3 = F)P(u_2 = Tr|w_2 = F)P(w_3 = F|w_2 = F)P(w_2 = F|w_1 = S)P(w_1 = S)}{P(u_3 = Fl)P(u_2 = Tr)P(w_1 = S)} \\
& + \frac{P(u_3 = Fl|w_3 = F)P(u_2 = Tr|w_2 = R)P(w_3 = F|w_2 = R)P(w_2 = R|w_1 = S)P(w_1 = S)}{P(u_3 = Fl)P(u_2 = Tr)P(w_1 = S)} \\
& + \frac{P(u_3 = Fl|w_3 = F)P(u_2 = Tr|w_2 = S)P(w_3 = F|w_2 = S)P(w_2 = S|w_1 = S)P(w_1 = S)}{P(u_3 = Fl)P(u_2 = Tr)P(w_1 = S)}
\end{aligned}
$$

Solution 6

- We need to calculate $P(w_3 = F | w_1 = S, u_2 = Tr, u_3 = Fl) = ?$

$$
\begin{aligned}
& P(w_3 = F | w_1 = S, u_2 = Tr, u_3 = Fl) \\
= \quad & P(w_2 = F, w_3 = F | w_1 = S, u_2 = Tr, u_3 = Fl) \\
& + P(w_2 = R, w_3 = F | w_1 = S, u_2 = Tr, u_3 = Fl) \\
& + P(w_2 = S, w_3 = F | w_1 = S, u_2 = Tr, u_3 = Fl) \\
= \quad & \frac{P(u_3 = Fl | w_3 = F) P(u_2 = Tr | w_2 = F) P(w_3 = F | w_2 = F) P(w_2 = F | w_1 = S) P(w_1 = S)}{P(u_3 = Fl) P(u_2 = Tr) P(w_1 = S)} \\
+ \quad & \frac{P(u_3 = Fl | w_3 = F) P(u_2 = Tr | w_2 = R) P(w_3 = F | w_2 = R) P(w_2 = R | w_1 = S) P(w_1 = S)}{P(u_3 = Fl) P(u_2 = Tr) P(w_1 = S)} \\
+ \quad & \frac{P(u_3 = Fl | w_3 = F) P(u_2 = Tr | w_2 = S) P(w_3 = F | w_2 = S) P(w_2 = S | w_1 = S) P(w_1 = S)}{P(u_3 = Fl) P(u_2 = Tr) P(w_1 = S)}
\end{aligned}
$$

Solution 6 contd.

$$
\begin{aligned}
= \quad & \frac{P(u_3 = Fl|w_3 = F)P(u_2 = Tr|w_2 = F)P(w_3 = F|w_2 = F)P(w_2 = F|w_1 = S))}{P(u_3 = Fl)P(u_2 = Tr)} \\
+ \quad & \frac{P(u_3 = Fl|w_3 = F)P(u_2 = Tr|w_2 = R)P(w_3 = F|w_2 = R)P(w_2 = R|w_1 = S)}{P(u_3 = Fl)P(u_2 = Tr)} \\
+ \quad & \frac{P(u_3 = Fl|w_3 = F)P(u_2 = Tr|w_2 = S)P(w_3 = F|w_2 = S)P(w_2 = S|w_1 = S)}{P(u_3 = Fl)P(u_2 = Tr)} \\
= \quad & \frac{0.7 \times 0.3 \times 0.5 \times 0.15}{0.5 \times 0.5} + \frac{0.7 \times 0.8 \times 0.2 \times 0.05}{0.5 \times 0.5} + \frac{0.7 \times 0.1 \times 0.15 \times 0.8}{0.5 \times 0.5} \\
= \quad & 0.119
\end{aligned}
$$

Thank You!