Hidden Markov Models

1 Markov Models

Let's talk about the weather. Let's say in Graz, there are three types of weather: sunny , rainy , and foggy . Let's assume for the moment that the weather lasts all day, i.e., it doesn't change from rainy to sunny in the middle of the day.

Weather *prediction* is about trying to guess what the weather will be like tomorrow based on the observations of the weather in the past (the *history*). Let's set up a *statistical model* for weather prediction: We collect statistics on what the weather q_n is like today (on day n) depending on what the weather was like yesterday q_{n-1}, the day before q_{n-2}, and so forth. We want to find the following conditional probabilities

$$P(q_n|q_{n-1}, q_{n-2}, ..., q_1), \tag{1}$$

meaning, the probability of the unknown weather at day n, $q_n \in \{\text{☀}, \text{🌧}, \text{☁}\}$, depending on the (known) weather $q_{n-1}, q_{n-2}, \ldots$ of the past days.

Using the probability in eq. 1, we can make probabilistic predictions of the type of weather for tomorrow and the next days using the observations of the weather history. For example, if we knew that the weather for the past three days was $\{\text{☀}, \text{☀}, \text{☁}\}$ in chronological order, the probability that tomorrow would be 🌧 is given by:

$$P(q_4 = \text{🌧}|q_3 = \text{☁}, q_2 = \text{☀}, q_1 = \text{☀}). \tag{2}$$

This probability could be inferred from the relative frequency (the *statistics*) of past observations of weather sequences $\{\text{☀}, \text{☀}, \text{☁}, \text{🌧}\}$.

Here's one problem: the larger n is, the more observations we must collect. Suppose that $n = 6$, then we have to collect statistics for $3^{(6-1)} = 243$ past histories. Therefore, we will make a simplifying assumption, called the Markov assumption:

For a sequence $\{q_1, q_2, ..., q_n\}$:

$$P(q_n|q_{n-1}, q_{n-2}, ..., q_1) = P(q_n|q_{n-1}). \tag{3}$$

This is called a *first-order Markov assumption*: we say that the probability of a certain observation at time n only depends on the observation q_{n-1} at time $n-1$. (A second-order Markov assumption would have the probability of an observation at time n depend on q_{n-1} and q_{n-2}. In general, when people talk about a Markov assumption, they usually mean the first-order Markov assumption.) A system for which eq. 3 is true is a *(first-order) Markov model*, and an output sequence $\{q_i\}$ of such a system is a *(first-order) Markov chain*.

We can also express the probability of a certain sequence $\{q_1, q_2, \ldots, q_n\}$ (the joint probability of certain past and current observations) using the Markov assumption:[2]

$$P(q_1, ..., q_n) = \prod_{i=1}^{n} P(q_i|q_{i-1}). \tag{4}$$

The Markov assumption has a profound affect on the number of histories that we have to find statistics for – we now only need $3 \cdot 3 = 9$ numbers ($P(q_n|q_{n-1})$ for every possible combination of $q_n, q_{n-1} \in \{\text{☀}, \text{🌧}, \text{☁}\}$) to characterize the probabilities of all possible sequences. The Markov assumption may or may not be a valid assumption depending on the situation (in the case of weather, it's probably not valid), but it is often used to simplify modeling.

So let's arbitrarily pick some numbers for $P(q_{\text{tomorrow}}|q_{\text{today}})$, as given in table 1 (note, that – whatever the weather is today – there *certainly is some kind of weather* tomorrow, so the probabilities in every *row* of table 1 sum up to one).

Today's weather	Tomorrow's weather		
	☀	🌧	☁
☀	0.8	0.05	0.15
🌧	0.2	0.6	0.2
☁	0.2	0.3	0.5

Table 1: Probabilities $p(q_{n+1}|q_n)$ of tomorrow's weather based on today's weather

For first-order Markov models, we can use these probabilities to draw a probabilistic finite state automaton. For the weather domain, we would have three states, $S = \{\text{☀}, \text{🌧}, \text{☁}\}$, and every day there would be a possibility $p(q_n|q_{n-1})$ of a transition to a (possibly different) state according to the probabilities in table 1. Such an automaton would look like shown in figure 1.

[2] One question that comes to mind is "What is q_0?" In general, one can think of q_0 as the *start word*, so $P(q_1|q_0)$ is the probability that q_1 can start a sentence. We can also just multiply a *prior probability* of q_1 with the product of $\prod_{i=2}^{n} P(q_i|q_{i-1})$, it's just a matter of definitions.

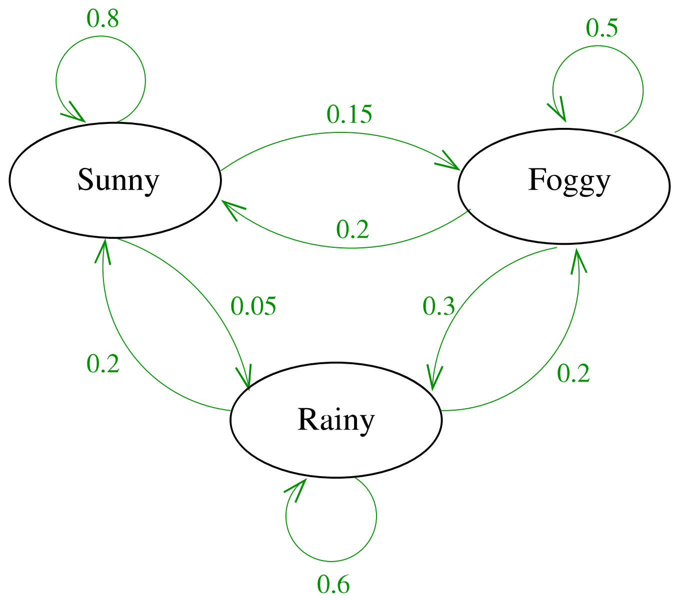

Figure 1: Markov model for the Graz weather with state transition probabilities according to table 1

1.0.1 Examples

1. Given that today the weather is ☀, what's the probability that tomorrow is ☀ and the day after is 🌧?
   Using the Markov assumption and the probabilities in table 1, this translates into:

$$\begin{aligned}
P(q_2 = \text{☀}, q_3 = \text{🌧} | q_1 = \text{☀}) &= P(q_3 = \text{🌧} | q_2 = \text{☀}, q_1 = \text{☀}) \cdot P(q_2 = \text{☀} | q_1 = \text{☀}) \\
&= P(q_3 = \text{🌧} | q_2 = \text{☀}) \cdot P(q_2 = \text{☀} | q_1 = \text{☀}) \qquad \text{(Markov assumption)} \\
&= 0.05 \cdot 0.8 \\
&= 0.04
\end{aligned}$$

   You can also think about this as moving through the automaton (figure 1), multiplying the probabilities along the path you go.

2. Assume, the weather yesterday was $q_1 = \text{🌧}$, and today it is $q_2 = \text{🌫}$, what is the probability that tomorrow it will be $q_3 = \text{☀}$?

$$\begin{aligned}
P(q_3 = \text{☀} | q_2 = \text{🌫}, q_1 = \text{🌧}) &= P(q_3 = \text{☀} | q_2 = \text{🌫}) \qquad \text{(Markov assumption)} \\
&= 0.2.
\end{aligned}$$

3. Given that the weather today is $q_1 = \text{🌫}$, what is the probability that it will be 🌧 two days from now: $q_3 = \text{🌧}$. (Hint: There are several ways to get from 🌫 today to 🌧 two days from now. You have to sum over these paths.)

2 Hidden Markov Models (HMMs)

So far we heard of the Markov assumption and Markov models. So, what is a *Hidden Markov Model*? Well, suppose you were locked in a room for several days, and you were asked about the weather outside. The only piece of evidence you have is whether the person who comes into the room bringing your daily meal is carrying an umbrella (☂) or not (🚫☂).

Let's suppose the probabilities shown in table 2: The probability that your caretaker carries an umbrella is 0.1 if the weather is sunny, 0.8 if it is actually raining, and 0.3 if it is foggy.

The equation for the weather Markov process before you were locked in the room was (eq. 4):

$$P(q_1, ..., q_n) = \prod_{i=1}^{n} P(q_i | q_{i-1}).$$

Weather	Probability of umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Table 2: Probability $P(x_i|q_i)$ of carrying an umbrella ($x_i = \text{true}$) based on the weather q_i on some day i

However, the actual weather is *hidden* from you. Finding the probability of a certain weather $q_i \in \{\text{☀}, \text{🌧}, \text{🌫}\}$ can only be based on the observation x_i, with $x_i = \text{☂}$, if your caretaker brought an umbrella on day i, and $x_i = \text{🚫☂}$ if the caretaker did not bring an umbrella. This conditional probability $P(q_i|x_i)$ can be rewritten according to Bayes' rule:

$$P(q_i|x_i) = \frac{P(x_i|q_i)P(q_i)}{P(x_i)},$$

or, for n days, and weather sequence $Q = \{q_1, \ldots, q_n\}$, as well as 'umbrella sequence' $X = \{x_1, , \ldots, x_n\}$

$$P(q_1, \ldots, q_n|x_1, \ldots, x_n) = \frac{P(x_1, \ldots, x_n|q_1, \ldots, q_n)P(q_1, \ldots, q_n)}{P(x_1, \ldots, x_n)},$$

using the probability $P(q_1, \ldots, q_n)$ of a Markov weather sequence from above, and the probability $P(x_1, \ldots, x_n)$ of seeing a particular sequence of umbrella events (e.g., $\{\text{☂}, \text{🚫☂}, \text{☂}\}$). The probability $P(x_1, \ldots, x_n|q_1, \ldots, q_n)$ can be estimated as $\prod_{i=1}^{n} P(x_i|q_i)$, if you assume that, for all i, the q_i, x_i are independent of all x_j and q_j, for all $j \neq i$.

We want to draw conclusions from our observations (if the persons carries an umbrella or not) about the weather outside. We can therefore omit the probability of seeing an umbrella $P(x_1, \ldots, x_n)$ as it is independent of the weather, that we like to predict. We get a measure for the probability, which is proportional to the probability, and which we will refer as the likelihood L.

$$\begin{aligned} &P(q_1, \ldots, q_n|x_1, \ldots, x_n) \propto \\ &L(q_1, \ldots, q_n|x_1, \ldots, x_n) = P(x_1, \ldots, x_n|q_1, \ldots, q_n) \cdot P(q_1, \ldots, q_n) \end{aligned} \tag{5}$$

With our (first order) Markov assumption it turns to:

$$\begin{aligned} &P(q_1, \ldots, q_n|x_1, \ldots, x_n) \propto \\ &L(q_1, \ldots, q_n|x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|q_i) \cdot \prod_{i=1}^{n} P(q_i|q_{i-1}) \end{aligned} \tag{6}$$

2.0.1 Examples

1. Suppose the day you were locked in it was sunny. The next day, the caretaker carried an umbrella into the room. You would like to know, what the weather was like on this second day.

   First we calculate the likelihood for the second day to be sunny:

   $$\begin{aligned} L(q_2 = \text{☀}|q_1 = \text{☀}, x_2 = \text{☂}) &= P(x_2 = \text{☂}|q_2 = \text{☀}) \cdot P(q_2 = \text{☀}|q_1 = \text{☀}) \\ &= 0.1 \cdot 0.8 = 0.08, \end{aligned}$$

   then for the second day to be rainy:

   $$\begin{aligned} L(q_2 = \text{🌧}|q_1 = \text{☀}, x_2 = \text{☂}) &= P(x_2 = \text{☂}|q_2 = \text{🌧}) \cdot P(q_2 = \text{🌧}|q_1 = \text{☀}) \\ &= 0.8 \cdot 0.05 = 0.04, \end{aligned}$$

   and finally for the second day to be foggy:

   $$\begin{aligned} L(q_2 = \text{🌫}|q_1 = \text{☀}, x_2 = \text{☂}) &= P(x_2 = \text{☂}|q_2 = \text{🌫}) \cdot P(q_2 = \text{🌫}|q_1 = \text{☀}) \\ &= 0.3 \cdot 0.15 = 0.045. \end{aligned}$$

   Thus, although the caretaker did carry an umbrella, it is most likely that on the second day the weather was sunny.

2. Suppose you do not know how the weather was when your were locked in. The following three days the caretaker always comes without an umbrella. Calculate the likelihood for the weather on these three days to have been $\{q_1 = , q_2 = , q_3 = \}$. As you do not know how the weather is on the first day, you assume the 3 weather situations are equi-probable on this day (cf. footnote on page 2), and the *prior probability* for sun on day one is therefore $P(q_1 = |q_0) = P(q_1 =) = 1/3$.

$$
\begin{aligned}
L(q_1 = , q_2 = , q_3 = | x_1 = , x_2 = , x_3 =) = \\
P(x_1 = |q_1 =) \cdot P(x_2 = |q_2 =) \cdot P(x_3 = |q_3 =) \cdot \\
P(q_1 =) \cdot P(q_2 = |q_1 =) \cdot P(q_3 = |q_2 =) = \\
0.9 \cdot 0.7 \cdot 0.9 \cdot 1/3 \cdot 0.15 \cdot 0.2 = 0.0057
\end{aligned}
\tag{7}
$$