Section 5
Markov Chains

As an application of matrix multiplication unrelated to systems of equations, let's consider Markov chains. But first, a quick reminder of some probability basics.

Suppose the probability of event E_1 is p_1 and the probability of event E_2 is p_2. For example, if a die is rolled, event E_1 could be rolling a 3; the probability of that would be $p_1 = 1/6$.

1. If E_1 and E_2 are independent events (so whether or not one event occurs does not depend on whether or not the other event occurs), then the probability that both E_1 *and* E_2 occur is $p_1 p_2$.

   Example
   A die is rolled, so the probability of rolling a 3 is 1/6. A coin is flipped, so the probability of getting a heads is 1/2. These two events are independent, so the probability of rolling a 3 and getting a heads is $\frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$.

   $\odot$

2. If E_1 and E_2 are mutually exclusive (so they can't both occur), then the probability that either E_1 *or* E_2 occurs is $p_1 + p_2$.

   Example
   A die is rolled, so the probability of rolling a 3 is $p_1 = 1/6$ and the probability of rolling an even number is 3/6. These two events are mutually exclusive, so the probability of rolling a 3 or an even number is $\frac{1}{6} + \frac{3}{6} = \frac{4}{6}$.

   $\odot$

Before we talk about Markov chains in general, let's start with an example.

Example
Every day, it either rains or it doesn't rain.

One day, the probability that it rains is 0.6 and so the probability that it doesn't rain is 0.4. Since these are all the possibilities, these probabilities need to add up to 1. We can store these values as a ***probability vector***:

$$\begin{array}{cc} R & NR \\ (0.6 & 0.4\,) \end{array}$$

(A probability vector is a vector of non-negative numbers which add to one. In this case, our probability vector is a row vector.) This tells us at a glance the probability of rain (R) and the probability of no rain (NR).

It's possible (likely, even) that the probability of rain one day depends on whether or not there was rain the previous day. If it rained the previous day, We might get the above probability vector

$$\begin{array}{ccc} & R & NR \\ R & (0.6 & 0.4\,) \end{array}.$$

Note that we are assuming that the probability of rain one day only depends on whether or not it rained the previous day, and nothing else.

If it didn't rain the previous day, we might get the probability vector

$$\begin{array}{cc} NR & \\ \begin{pmatrix} 0.2 \\ 0.8 \end{pmatrix} & \begin{array}{c} R \\ NR \end{array} \end{array} \cdot \begin{array}{ccc} & R & NR \\ NR & (0.2 & 0.8\,) \end{array}$$

The complete situation can be described by making the probability vectors the rows of a matrix:

$$\begin{array}{cc} & \begin{array}{cc} R & NR \end{array} \\ \begin{array}{c} NR \\ R \end{array} & \begin{pmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{pmatrix} \end{array}$$

Note that the rows and columns are labeled in the same order. This is an example of a stochastic matrix, a square matrix where each column is a probability vector.

Now suppose that one day, somehow, the probability of rain is given by the probability vector

$$\mathbf{q} = \begin{array}{cc} R & NR \\ (0.25 & 0.75) \end{array}.$$

What is the probability of rain the next day? There are two possibilities for rain the next day;

1. It rains today, and it rains tomorrow.

2. It doesn't rain today, and it rains tomorrow.

The probabilities of each of these possibilities are

1. The probability of rain today times the probability of rain being followed by rain; 0.25×0.6,

2. The probability of no rain today times the probability of no rain being followed by rain; 0.75×0.2.

Altogether the probability of rain tomorrow will be

$$0.25 \times 0.6 + 0.75 \times 0.2 = 0.3.$$

Note that this is the product

$$\begin{pmatrix} 0.25 & 0.75 \end{pmatrix} \cdot \begin{pmatrix} 0.6 \\ 0.2 \end{pmatrix}$$

which is the rain column of the transition matrix times today's probability vector. Similarly, the probability of no rain tomorrow will be the no rain column of the transition matrix times today's probability vector,

$$\begin{pmatrix} 0.25 & 0.75 \end{pmatrix} \cdot \begin{pmatrix} 0.4 \\ 0.8 \end{pmatrix} = 0.7.$$

Altogether, tomorrow's probability vector will be

$$\begin{pmatrix} 0.25 & 0.75 \end{pmatrix} \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.3 & 0.7 \end{pmatrix},$$

which is today's probability vector times the transition matrix.

$\odot$

This situation is an example of a Markov chain. More generally, we have the following definition.

Definition

A ***Markov Chain*** or ***Markov Process*** is a system with the following properties:

1. There are a finite number of possible states: $\{E_1, E_2, \ldots, E_n\}$.
2. Changes occur in discrete steps.
3. When in state E_i, the probability p_{ij} of going to state E_j depends only on E_i and E_j.

$\bigcirc$

Each state in a Markov chain will have a probability vector, giving the probability of moving to a different state in the next step.

$$\begin{array}{cc} & \begin{array}{cccc} E_1 & E_2 & \ldots & E_n \end{array} \\ E_i & \begin{pmatrix} p_{1i} & p_{2i} & \vdots & p_{ni} \end{pmatrix} \end{array}$$

It's convenient to display these probability vectors as a stochastic matrix:

$$\begin{array}{cc} & \begin{array}{cccccc} E_1 & E_2 & E_3 & \ldots & E_n \end{array} \\ \begin{array}{c} E_1 \\ E_2 \\ E_3 \\ \vdots \\ E_n \end{array} & \begin{pmatrix} p_{11} & p_{12} & p_{13} & \ldots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \ldots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \ldots & p_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \ldots & p_{nn} \end{pmatrix} \end{array}$$

This matrix is called the ***transition matrix*** for the Markov chain; it stores all the probabilities involved in going from one state to another.

Proposition

Suppose that a Markov process with transition matrix P has an initial probability vector $\mathbf{q}$. Then the probability vector for the state after one step is $\mathbf{q}P$.

$\square$

Example

Suppose a Markov process has an initial probability vector of $\mathbf{q} = \begin{pmatrix} .25 & .5 & .25 \end{pmatrix}$ and a transition matrix of

$$P = \begin{pmatrix} 0.4 & 0.4 & 0.2 \\ 0.6 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}$$

the probability vector for the second step is

$$\mathbf{q}P = \begin{pmatrix} 0.45 & 0.25 & 0.3 \end{pmatrix}$$

$\odot$

Continuing, since the probability vector after one step is $\mathbf{q}P$, the probability vector after two steps will be $(\mathbf{q}P)P = \mathbf{q}P^2$, etc. Each step introduces another power of P.

Proposition

If a Markov process has transition matrix P, then the transition matrix corresponding to k steps is P^k.

$\square$

Example

Using our example with transition matrix

$$P = \begin{array}{c} \\ M \\ W \\ S \end{array} \!\!\begin{array}{c} \begin{array}{ccc} M & W & S \end{array} \\ \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.1 & 0.7 & 0.2 \end{pmatrix} \end{array}$$

the transition matrix for two steps is

$$P^2 = \begin{array}{c} \\ M \\ W \\ S \end{array} \!\!\begin{array}{c} \begin{array}{ccc} M & W & S \end{array} \\ \begin{pmatrix} 0.21 & 0.50 & 0.29 \\ 0.23 & 0.42 & 0.35 \\ 0.32 & 0.38 & 0.3 \end{pmatrix} \end{array}.$$

So, for example, the probability of drinking soda for breakfast two days after having water for breakfast is 0.35.

$\odot$