Conversation 36: Applications of Eigenvectors to Markov Chains

Winfried Just
Department of Mathematics, Ohio University

MATH3200: Applied Linear Algebra

Remember Waldo?

Denny: It's kind of curious how this linear algebra stuff pops up in unexpected places. Remember this guy Waldo that I told you about? I called him the other day and told him that we had figured out what he was after with his random web surfing and lists of websites that he hit upon.

Frank: The PageRank guy? Yeah, I recall that we had figured out that at Google they do this random surfing big time to tabulate frequency of hits and rank websites by popularity.

Denny: That's what I told him. And he said: "No. We do it differently these days." You'd never guess what he told me next.

Alice: They calculate left eigenvectors of the transition probability matrix.

Remember Markov chains?

Denny: That's exactly what he said! I thought you'd never guess.

Frank: Never say never, especially when it comes to Alice.

Alice: Thank you Frank! I may take you up on that one.

Frank: What are you talking about, Alice?!? But I have forgotten all about these Markov chains. Can you remind us how they work?

Alice: They are stochastic processes, that is, mathematical models for studying situations in which the sate of a system changes somewhat unpredictably over time.

Theo: Time is assumed to proceed in discrete steps $t = 0, 1, 2, \ldots$ At each time t the process can only be in one of several states that are numbered $1, \ldots, n$. The probability of being in a given state at time $t + 1$ depends only on the state at time t.

Review: Markov chains for weather.com light

Cindy: Like in weather.com light. One time step lasted one day and the states were State 1: sunny day, State 2: rainy day.

Theo: Right. And for each state we had certain transition probabilities to the state of the weather on the next day:

- p_{11} is the probability that a sunny day is followed by another sunny day.
- p_{12} is the probability that a sunny day is followed by a rainy day.
- p_{21} is the probability that a rainy day is followed by a sunny day.
- p_{22} is the probability that a rainy day is followed by another rainy day.

Bob: I remember that we organized them into a matrix

of transition probabilities $\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$

An eigenvector of the matrix P

Theo: Exactly! Each Markov chain is characterized by such a matrix $\mathbf{P}$. It must be a *stochastic matrix.*

Denny: What was that again?

Bob: It means that all the elements of $\mathbf{P}$ are probabilities and each row adds up to 1.

Theo: For each stochastic matrix $\mathbf{P}$ the vector $\vec{\mathbf{1}} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

is an eigenvector of $\mathbf{P}$.

Question C36.1: Why would this be true, and what is the eigenvalue of $\vec{\mathbf{1}}$?

Theo: Because $\mathbf{P}\vec{\mathbf{1}}$ is the vector of sums of all rows of $\mathbf{P}$. So if $\mathbf{P}$ is stochastic, then $\mathbf{P}\vec{\mathbf{1}} = \vec{\mathbf{1}}$, and we can see that $\vec{\mathbf{1}}$ is an eigenvector of $\mathbf{P}$ with eigenvalue $\lambda^* = 1$.

How about left eigenvectors of the matrix P?

Denny: Why did you write λ^* instead of λ, Theo?

Theo: Because this is a common notation for the so-called *leading eigenvalue* of a matrix, which is an eigenvalue λ^* so that $|\lambda| \leq |\lambda^*|$ for all other eigenvalues of this matrix.

There is a famous theorem, called the *Perron-Frobenius Theorem,* which implies that for every stochastic matrix $\lambda^* = 1$ will be a leading eigenvalue. I can show you—

Alice: Maybe not now. Let's talk about left eigenvectors of transition probability matrices $\mathbf{P}$. They are more interesting.

Frank: Right! The last time you promised to show us some application of left eigenvectors. So what about them?

Alice: As Theo has just shown us, when $\mathbf{P}$ is a transition probability matrix of a Markov chain, then $\lambda^* = 1$ must be an eigenvalue of $\mathbf{P}$.

Alice: Then $\lambda^* = 1$ must also be an eigenvalue of $\mathbf{P}^T$, so that there exists an eigenvector of $\mathbf{P}^T$ with eigenvalue $\lambda^* = 1$.

When we take the transpose $\vec{\mathbf{x}}$ of such an eigenvector of $\mathbf{P}^T$, then we obtain a left eigenvector of $\mathbf{P}$ that satisfies $\vec{\mathbf{x}}\mathbf{P} = \vec{\mathbf{x}}$.

Theo: And the Perron-Frobenius Theorem implies that we can always find such a vector $\vec{\mathbf{x}}$ that is a probability distribution.

Denny: Can you remind us what these probability distributions were, Theo?

Theo: They are row vectors $\vec{\mathbf{x}} = [x_1, \dots, x_n]$ such that each component x_i is a probability, that is, $0 \leq x_i \leq 1$, and $x_1 + \cdots + x_n = 1$.

Probability distributions

Alice: We can think of the probability distribution $\vec{\mathbf{x}}(t) = [x_1(t), \dots, x_n(t)]$ as our estimates of the probabilities $x_i(t)$ that state i will be observed at time t.

Cindy: Like in the weather.com example. When $\vec{\mathbf{x}}(3) = [0.2, 0.8]$, there would be a 20% chance that day number 3 will be a sunny day and an 80% chance that day number 3 is a rainy day.

Theo: And recall that the probability distribution $\vec{\mathbf{x}}(t+1)$ on the next day is always given by the matrix product $\vec{\mathbf{x}}(t+1) = \vec{\mathbf{x}}(t)\mathbf{P}$.

Question C36.2: If our estimate $\vec{\mathbf{x}}(t)$ of the probabilities on day t happens to be a left eigenvector of $\mathbf{P}$ with eigenvalue 1, what does this imply about our estimates $\vec{\mathbf{x}}(t+1)$?

Theo: Then $\vec{\mathbf{x}}(t+1) = \vec{\mathbf{x}}(t)\mathbf{P} = 1\vec{\mathbf{x}}(t) = \vec{\mathbf{x}}(t)$. So our estimates of the probability distribution on day $t+1$ would be the same as for day t. Such a vector $\vec{\mathbf{x}}(t)$ where the probability distribution does not change is called a *stationary distribution* of the Markov chain.

Stationary states and long-range forecasts

Bob: Let me see whether I get this straight: If $\vec{\mathbf{x}}(t)$ is such a stationary distribution, would we then also have $\vec{\mathbf{x}}(t+2) = \vec{\mathbf{x}}(t)$?

Alice: Yes you would. Notice that in this case
$\vec{\mathbf{x}}(t+2) = \vec{\mathbf{x}}(t)\mathbf{P}^2 = (\vec{\mathbf{x}}(t)\mathbf{P})\mathbf{P} = \vec{\mathbf{x}}(t)\mathbf{P} = \vec{\mathbf{x}}(t)$.

Cindy: Then also $\vec{\mathbf{x}}(t+3) = \vec{\mathbf{x}}(t)$, by the same argument, right?

Alice: Exactly! In fact, we would have $\vec{\mathbf{x}}(t+k) = \vec{\mathbf{x}}(t)$ for all $k > 0$, so that the estimated distribution does never change in the future when we start in a stationary distribution.

Frank: Are you saying, Alice, that then the weather would not change in the future? This doesn't make sense.

Alice: No. The actual weather will change. But our uncertainty about the weather in the future, which is given by $\vec{\mathbf{x}}(t)$ will remain exactly the same if $\vec{\mathbf{x}}(t)$ is a stationary distribution. You can think of such $\vec{\mathbf{x}}(t)$ as giving you long-range probabilities for each state.

How to interpret stationary distributions?

Question C36.3: For our weather.com example with
$\mathbf{P} = \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix}$
what would a stationary distribution be?

Alice: Here $\vec{\mathbf{x}} = [1/3, 2/3]$ is the only stationary distribution.

Cindy: Isn't this he same distribution that I got when I made a five-day forecast with this Markov chain by calculating $[0, 1]\mathbf{P}^5$?

Alice: Not exactly the same, but very close. You got the same numbers with an accuracy of 4 digits after the decimal point.

Theo: In fact, for most Markov chains, regardless of the initial state $\vec{\mathbf{x}}(0)$, the long-range estimates $\vec{\mathbf{x}}(t)$ will very quickly approach the stationary distribution that gives the average proportions x_i of observing state i in a long sequence of time steps.

We will explore in Module 71 when and how this works.

Denny: What does this have to do with Waldo's work at Google?

Alice: You told us that when in college, Waldo was randomly following links and keeping checklists of how often he would hit a given web page. So if he were to do this long enough, these checklists would give him estimates of the frequencies of hits for any given page i.

Frank: Come on, Alice! He would need to do this billions and billions of times to get reasonable estimates!

Alice: Exactly! This would be too tedious, even for a company that has the computational resources of Google. But Google knows the transition transition probability matrix **P** for this huge Markov chain and could get the same result by computing the stationary distribution, the left eigenvector of **P** with eigenvalue $\lambda^* = 1$.

How is this related to popularity?

Denny: Cool! But I still don't see how these x_i in the stationary distribution are related to "popularity."

Theo: As Alice has told us, a stationary distribution will give us, for each website i, the average proportion x_i of times this website is visited in a long session of random surfing, the kind of random surfing that Waldo did back in his dorm room. The higher this proportion x_i, the more often website number i gets visited.

Cindy: Oh, I see! The most "popular" websites are the ones that are most often visited!

Frank: I still don't buy this. Neither Waldo nor anybody else really clicks randomly on billions of links.

Waldo approximates the surfing of billions of users

Alice: Yes, but we can assume that billions of users would more or less randomly follow links for some time. As Cindy has explained to us, the frequency of visits by these billions of users would give us a measure the "popularity" of web page number i.

The proportions x_i of one very, very long session of random surfing would give us some reasonable approximations of these frequencies. Simulating such very, very long sessions is not feasible, even on a powerful computer, but we can obtain them by calculating the stationary distributions as a left eigenvector of the Markov chain.

Frank: Sounds plausible. But how about this "teleporting" that Waldo supposedly did once in a while? Those billions of users don't have a directory of all URL's.

Alice: No. But they will sometimes open random URLs that have been recommended to them by friends. This would not be entirely unlike teleporting.

Denny: That's it! I could recommend some really cool ones!

Take-home message

Eigenvectors and eigenvalues have many important applications.

Transition probability matrices of Markov chains always have an eigenvalue $\lambda = 1$. Moreover, they have at least one left eigenvector with eigenvalue $\lambda = 1$ that is a probability distribution. It is called a *stationary distribution* of the Markov chain.

When $\vec{\mathbf{x}}(t)$ is a stationary distribution, then the distribution at the next time step, and at all subsequent time steps, will exactly be the same:

$\vec{\mathbf{x}}(t) = \vec{\mathbf{x}}(t+1) = \cdots = \vec{\mathbf{x}}(t+k) = \dots$

Stationary distributions also represent expected frequencies of the states over many time steps.