Stochastic Matrices

Today we'll discuss a fun application of eigenvalues: Google's PageRank algorithm. It's an instance of a stochastic process, which is a kind of difference equation representing probabilities.

Running Example: Red Box:

Pretend there are 3 RedBox kiosks in Durham. Assume that everyone who rents Prognosis Negative today will return it tomorrow.

The probability that someone renting from kiosk i will return to kiosk j is given in this table:

Returning to \ Renting from	1	2	3
1	30%	40%	50%
2	30%	40%	30%
3	40%	20%	20%

so e.g. someone renting from kiosk 3 has a 50% chance of returning to kiosk 1.

Question: where will all the copies of Prognosis Negative eventually end up?

We can represent this as a difference equation!

x_k = # movies in kiosk 1
y_k = # movies in kiosk 2
z_k = # movies in kiosk 3

on day k

The rules say:

$$x_{k+1} = .3x_k + .4y_k + .5z_k$$
$$y_{k+1} = .3x_k + .4y_k + .3z_k$$
$$z_{k+1} = .4x_k + .2y_k + .2z_k$$

"tomorrow, kiosk 1 will have 30% of the movies from kiosk 1, 40% from kiosk 2, & 50% from kiosk 3"

As a vector equation:

$$v_k = \begin{pmatrix} x_k \\ y_k \\ z_k \end{pmatrix} \qquad A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix} \qquad v_{k+1} = Av_k$$

NB: the columns of A sum to 1 because we're assuming every movie has a 100% chance of being returned somewhere.

This means that the total # movies does not change: there are the same # movies on day $k+1$ as on day k.

sums to 1

$$\begin{aligned} &(x_{k+1} = .3x_k + .4y_k + .5z_k) \\ +&(y_{k+1} = .3x_k + .4y_k + .3z_k) \\ +&(z_{k+1} = .4x_k + .2y_k + .2z_k) \\ \hline &x_{k+1} + y_{k+1} + z_{k+1} = 1x_k + 1y_k + 1z_k \end{aligned}$$

Def:

- A square matrix is **stochastic** if its entries are ≥ 0 and the entries in each column sum to 1.
- A stochastic matrix is **positive** if its entries are all > 0 (ie, nonzero).

Eg: positive stochastic

$$\begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$$

stochastic

$$\begin{pmatrix} .6 & .4 & .5 \\ 0 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$$

not stochastic

$$\begin{pmatrix} .6 & .4 & .5 \\ -.1 & .4 & .3 \\ .5 & .2 & .2 \end{pmatrix}$$

not stochastic

$$\begin{pmatrix} .3 & .4 & .5 \\ .4 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$$

NB: The columns sum to 1 $\iff A^T\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$:

$$A=\begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix} \qquad A^T=\begin{pmatrix} .3 & .3 & .4 \\ .4 & .4 & .2 \\ .5 & .3 & .2 \end{pmatrix}$$

sum of col 1

$$A^T\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix} .3+.3+.4 \\ .4+.4+.2 \\ .5+.3+.2 \end{pmatrix}=\begin{pmatrix}1\\1\\1\end{pmatrix}$$

In particular, $\begin{pmatrix}1\\1\\1\end{pmatrix}$ is an eigenvector of A^T with eigenvalue 1. You showed on the HW that A and A^T have the same eigenvalues (same characteristic polynomial), so:

Consequence: $\boxed{\text{1 is an eigenvalue of any stochastic matrix}}$

What's better is that 1 is the dominant eigenvalue.

Fact: If λ is an eigenvalue of a stochastic matrix, then $|\lambda| \le 1$.

Proof: As above, λ is also an eigenvalue of A^T. Let v be an eigenvector, so $A^T v = \lambda v$.

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \implies A^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

$$v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \end{pmatrix} = \lambda v = A^T v = \begin{pmatrix} a_{11}x_1 + a_{21}x_2 + a_{31}x_3 \\ a_{12}x_1 + a_{22}x_2 + a_{32}x_3 \\ a_{13}x_1 + a_{23}x_2 + a_{33}x_3 \end{pmatrix}$$

Suppose $|x_1| \ge |x_2|$ and $|x_1| \ge |x_3|$

(choose the coordinate of v with maximal abs. value)

Then the 1^{st} coordinate of $\lambda v = A^T v$ is

$$\lambda x_1 = a_{11}x_1 + a_{21}x_2 + a_{31}x_3$$

$$\implies |\lambda| \cdot |x_1| = |a_{11}x_1 + a_{21}x_2 + a_{31}x_3|$$

$$\le \overset{\ge 0}{a_{11}} \cdot |x_1| + \overset{\ge 0}{a_{21}} \cdot \overset{\le |x_1|}{|x_2|} + \overset{\ge 0}{a_{31}} \cdot \overset{\le |x_1|}{|x_3|}$$

$$\le a_{11} \cdot |x_1| + a_{21} \cdot |x_1| + a_{31} \cdot |x_1|$$

$$= (a_{11} + a_{21} + a_{31})|x_1| = |x_1|$$

$$\implies |\lambda| \le 1$$

//

(harder to prove)

Better Fact: If $\lambda \ne 1$ is an eigenvalue of a positive stochastic matrix, then $|\lambda| < 1$.

So in this case, $\lambda = 1$ is really the dominant eigenvalue.

Running Example, Cont'd:

The RedBox matrix has characteristic polynomial

$$p(\lambda) = -\lambda^3 + .9\lambda^2 + 0.12\lambda - 0.02$$
$$= -(\lambda-1)(\lambda+0.2)(\lambda-0.1)$$

so the eigenvalues are $1, -0.2, 0.1$

and $1 > |-0.2|, \quad 1 > |0.1|$ ✓

In this case, there are 3 eigenvalues, so the matrix is diagonalizable. We compute some eigenvectors:

$$1: \; w_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix} \qquad -0.2: \; w_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \qquad 0.1: \; w_3 = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}$$

Suppose we started with $v_0 = \begin{pmatrix} 48 \\ 36 \\ 42 \end{pmatrix}$ movies.

Expand in the eigenbasis:

$$v_0 = x_1 w_1 + x_2 w_2 + x_3 w_3 \overset{\text{solve}}{\rightsquigarrow} x_1 = 7 \quad x_2 = 3 \quad x_3 = 2$$

$$\Longrightarrow v_0 = 7w_1 + 3w_2 + 2w_3$$

$$\Longrightarrow v_k = A^k v_0 = 1^k \cdot 7w_1 + (-0.2)^k \cdot 3w_2 + (0.1)^k \cdot 2w_3$$

$$\xrightarrow{k \to \infty} 7w_1 = \begin{pmatrix} 49 \\ 42 \\ 35 \end{pmatrix}$$

Observation 1: If $v_0 = x_1 w_1 + x_2 w_2 + x_3 w_3$

then $v_k = x_1 w_1 + (-0.2)^k x_2 w_2 + (0.1)^k x_3 w_3$

$$\xrightarrow{k \to \infty} x_1 w_1$$

So v_k converges to a 1-eigenvector (or 0 if $x=0$)

Observation 2: The total # movies doesn't change, so the sum of the entries in $x_1 w_1$ = the sum of the entries in v_0. This is a much easier way to solve for x_1!

Running Example, Cont'd: In our example, we had

$$v_0 = \begin{pmatrix} 48 \\ 36 \\ 42 \end{pmatrix} \rightsquigarrow 126 \text{ total movies.}$$

We chose $w_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}$ as our 1-eigenvector, so the sum of the entries in

$$x_1 w_1 = \begin{pmatrix} 7x_1 \\ 6x_1 \\ 5x_1 \end{pmatrix} \text{ is } (7+6+5)x_1 = 18x_1 \text{ movies.}$$

The total # movies doesn't change, so

$$126 = 18x_1 \implies x_1 = 7 \implies v_k \to 7x_1 = \begin{pmatrix} 49 \\ 42 \\ 35 \end{pmatrix} \checkmark$$

We didn't need to expand in the eigenbasis to figure this out!

NB: This calculation would've been easier if we'd chosen $w_1 = \frac{1}{18}\begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}$, because then the entries of w_1 sum to $1 \implies x_1$ is just the total # movies we started with.

Observation 3:

The coordinates of $w_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}$ are positive numbers.

They had better not be negative — otherwise one of the kiosks ends up with negative movies!

These observations hold for any positive stochastic matrix, even when it's not diagonalizable.

Perron-Frobenius Theorem: If A is a positive stochastic matrix, then there is a unique 1-eigenvector w with positive coordinates summing to 1.

If v_0 is any vector with coordinates summing to c, then

$$v_k = A^k v_0 \xrightarrow{k \to \infty} c \cdot w.$$

[DEMO]

Def: If A is a positive stochastic matrix, its unique 1-eigenvector with coordinates summing to 1 is its steady state.

This is relatively easy to compute!

How to Compute the Steady State of a Positive Stochastic Matrix:

(1) Find a nonzero 1-eigenvector (solve $(A - I_n)v = 0$)

(2) The steady state is $w = \frac{1}{\text{sum of coords of } v} \cdot v$.

Eg: The steady state of the RedBox matrix is

$$w = \frac{1}{7+6+5}\begin{pmatrix}7\\6\\5\end{pmatrix} = \frac{1}{18}\begin{pmatrix}7\\6\\5\end{pmatrix}.$$

Perron-Frobenius Theorem: Summary

Let A be a positive stochastic matrix.

- The 1-eigenspace of A is a line.
- There is a 1-eigenvector with positive entries.
  (divide by the sum of the coordinates ⇝)
- There is a unique 1-eigenvector with positive entries summing to 1.
- $|\lambda| < 1$ for all other eigenvalues, so 1 is the dominant eigenvalue.
- If v_0 is any vector then
  $$v_k = A^k v_0 \xrightarrow{k \to \infty} c \cdot w$$
- The scalar c is the sum of the coordinates of v_0. (total # moves doesn't change)

Google's PageRank

This is how Larry Page & Sergei Brin used eigenvalues to make it possible to find things on the Internet. They published the math behind their algorithm, so this is all for real.

Idea: Each web page has an "importance", or rank.

For any page P,

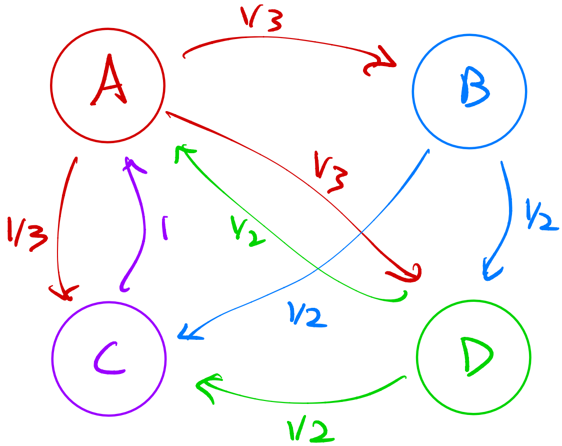
- $\text{rank}(P) > 0$
- If P links to n other pages $Q_1, Q_2, \ldots, Q_n$, then each Q_i inherits $\frac{1}{n} \cdot \text{rank}(P)$ for its rank.
  - → So if an important page links to P, then P is important too.
  - → Or, if a million unimportant pages link to P, then P is still important.
  - → But if only a few unimportant pages link to P, then P is not important.

Random Surfer Interpretation:

The random surfer sits at his computer all day, clicking links at random. It turns out that

$\text{rank}(P)$ = probability that he's visiting P at any given time.

Eg: Here's an Internet with 4 pages. Links are indicated by arrows.



- Page A has 3 links: passes $1/3$ of its rank to B C D.
- Page B has 2 links: passes $1/2$ of its rank to C D.
- Page C has 1 link: passes all of its rank to A.
- Page D has 2 links: passes $1/2$ of its rank to A C.

Let's write this using equations. Let a b c d denote the ranks of A B C D, respectively. Then:

$$
\begin{aligned}
a &= \phantom{\tfrac{1}{3}a + \tfrac{1}{2}b + {}} c + \tfrac{1}{2}d \\
b &= \tfrac{1}{3}a \\
c &= \tfrac{1}{3}a + \tfrac{1}{2}b \phantom{{}+c} + \tfrac{1}{2}d \\
d &= \tfrac{1}{3}a + \tfrac{1}{2}b
\end{aligned}
\quad \rightsquigarrow \quad
\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}
$$

This matrix is called the importance matrix.

Observation:

- The importance matrix is stochastic!

  For instance, the 1^{st} column sums to 1 because the first page gives $1/3$ of its rank to 3 pages ($1/3 \times 3 = 1$).

- The rank vector (a, b, c, d) is an eigenvector with eigenvalue 1

  (the \$25 billion eigenvector).

In this example, the 1-eigenspace is spanned by

$$w = \frac{1}{31}\begin{pmatrix} 12 \\ 4 \\ 9 \\ 6 \end{pmatrix} \implies a = \frac{12}{31} \quad b = \frac{4}{31} \quad c = \frac{9}{31} \quad d = \frac{6}{31}$$

(normalize so the sum is 1)

So A is your top hit!

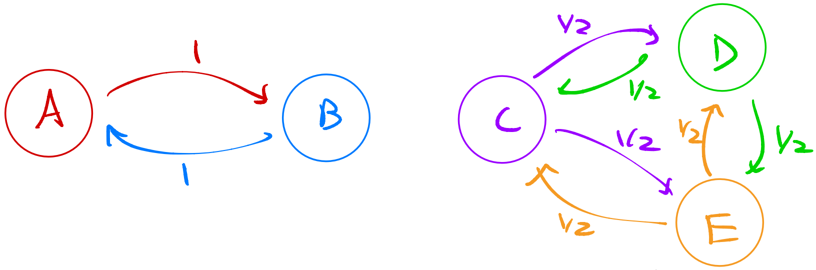
Random Surfer Interpretation: If the random surfer has probabilities (a, b, c, d) of being on page A, B, C, D, respectively, then after the next click he has probabilities

$$\begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

of being on each page. So the rank vector is the steady state for the random surfer $\leadsto$ his eventual probability of being on each page.

Observation: The importance matrix is stochastic, but not positive. Does this cause problems? Yes!

Eg (Disconnected Internet): Consider this Internet:



Its importance matrix is
$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

The 1-eigenspace has basis
$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

so the rank vector is not unique!

Eg: If a page has no links at all, then its column is all zero $\rightsquigarrow$ importance matrix is not stochastic!
(This is easy to fix — if a page has no links, pretend it links only to itself.)

Page & Brin's solution is as follows.

Fix a damping factor $p \in (0,1)$. (eg. $p = 0.15$)

Let A be the importance matrix and let

$$B = \frac{1}{N}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \qquad N = \#\text{ pages } (N \times N \text{ matrix})$$

Def: The **Google Matrix** is

$$G = (1-p)A + pB$$

Eg: In the disconnected internet example,

$$G = (1-p)\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} + p\begin{pmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}$$

Fact: The Google matrix is **positive stochastic**.

- Stochastic: The columns of $(1-p)A$ sum to $1-p$
  the columns of pB sum to p
  $\Rightarrow$ the columns of G sum to 1 ✓
- Positive: the entries of A are ≥ 0
  and the entries of B are > 0. ✓

Random Surfer Interpretation: With probability p, the random surfer navigates to a random page anywhere on the Internet; with probability $1-p$, he clicks a random link.

Def: The PageRank vector is the steady state of the Google matrix.

So a page's rank is just the value of its coordinate in the PageRank vector.

NB: "PageRank" is named after Larry Page.

This has been a discussion of what Google computes when you search for something. Of course, the real Internet has zillions of pages, and it's not clear how to compute an eigenvector of a zillion×zillion matrix. For Google, it's an industry secret!

One real-world algorithm is Arnoldi Iteration, which is good at finding eigenvectors of large matrices with lots of 0's. (See the webpage for a link to Wikipedia.)