Math 312

Markov chains, Google's PageRank algorithm

Jeff Jauregui

October 25, 2012

Random processes

- Goal: model a **random process** in which a system **transitions** from one **state** to another at discrete time steps.

Random processes

- Goal: model a **random process** in which a system **transitions** from one **state** to another at discrete time steps.
- At each time, say there are n states the system could be in.

Random processes

- Goal: model a **random process** in which a system **transitions** from one **state** to another at discrete time steps.
- At each time, say there are n states the system could be in.
- At time k, we model the system as a vector $\vec{x}_k \in \mathbb{R}^n$ (whose entries represent the probability of being in each of the n states).

Random processes

- Goal: model a **random process** in which a system **transitions** from one **state** to another at discrete time steps.
- At each time, say there are n states the system could be in.
- At time k, we model the system as a vector $\vec{x}_k \in \mathbb{R}^n$ (whose entries represent the probability of being in each of the n states).
- Here, $k = 0, 1, 2, \ldots$, and "initial state" is $\vec{x}_0$.

Random processes

- Goal: model a **random process** in which a system **transitions** from one **state** to another at discrete time steps.
- At each time, say there are n states the system could be in.
- At time k, we model the system as a vector $\vec{x}_k \in \mathbb{R}^n$ (whose entries represent the probability of being in each of the n states).
- Here, $k = 0, 1, 2, \ldots$, and "initial state" is $\vec{x}_0$.

Definition

A **probability vector** is a vector in $\mathbb{R}^n$ whose entries are nonnegative and sum to 1.

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$ (meaning 60% live in city, 40% live in suburbs).

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$ (meaning 60% live in city, 40% live in suburbs).
- Given: each year, 5% of city dwellers move to suburbs (the rest stay), and

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$ (meaning 60% live in city, 40% live in suburbs).
- Given: each year, 5% of city dwellers move to suburbs (the rest stay), and
- 3% of suburbanites move to city (the rest stay)

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$ (meaning 60% live in city, 40% live in suburbs).
- Given: each year, 5% of city dwellers move to suburbs (the rest stay), and
- 3% of suburbanites move to city (the rest stay)
- Let $\vec{x}_k = (c_k, s_k)$. We're told:

$$c_{k+1} = 0.95c_k + 0.03s_k$$
$$s_{k+1} = 0.05c_k + 0.97s_k$$

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$ (meaning 60% live in city, 40% live in suburbs).
- Given: each year, 5% of city dwellers move to suburbs (the rest stay), and
- 3% of suburbanites move to city (the rest stay)
- Let $\vec{x}_k = (c_k, s_k)$. We're told:

$$c_{k+1} = 0.95c_k + 0.03s_k$$
$$s_{k+1} = 0.05c_k + 0.97s_k$$

- i.e.

$$\vec{x}_{k+1} = M\vec{x}_k.$$

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$ (meaning 60% live in city, 40% live in suburbs).
- Given: each year, 5% of city dwellers move to suburbs (the rest stay), and
- 3% of suburbanites move to city (the rest stay)
- Let $\vec{x}_k = (c_k, s_k)$. We're told:

$$c_{k+1} = 0.95c_k + 0.03s_k$$
$$s_{k+1} = 0.05c_k + 0.97s_k$$

- i.e.

$$\vec{x}_{k+1} = M\vec{x}_k.$$

- $\vec{x}_1 = (0.58, 0.42)$, $\vec{x}_2 = (0.56, 0.44)$, $\vec{x}_{10} = (0.47, 0.53)$, etc.

Cities/suburbs

- Model a population in the city vs. suburbs. Say $\vec{x}_0 = (0.60, 0.40)$ (meaning 60% live in city, 40% live in suburbs).
- Given: each year, 5% of city dwellers move to suburbs (the rest stay), and
- 3% of suburbanites move to city (the rest stay)
- Let $\vec{x}_k = (c_k, s_k)$. We're told:

$$\begin{aligned} c_{k+1} &= 0.95c_k + 0.03s_k \\ s_{k+1} &= 0.05c_k + 0.97s_k \end{aligned}$$

- i.e.

$$\vec{x}_{k+1} = M\vec{x}_k.$$

- $\vec{x}_1 = (0.58, 0.42)$, $\vec{x}_2 = (0.56, 0.44)$, $\vec{x}_{10} = (0.47, 0.53)$, etc.
- For k large, $\vec{x}_k$ limits to $(0.375, 0.625)$.

Markov chains

Definition

A **Markov matrix** (or **stochastic matrix**) is a square matrix M whose columns are probability vectors.

Markov chains

Definition

A **Markov matrix** (or **stochastic matrix**) is a square matrix M whose columns are probability vectors.

Definition

A **Markov chain** is a sequence of probability vectors $\vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$ such that $\vec{x}_{k+1} = M\vec{x}_k$ for some Markov matrix M.

Markov chains

Definition

A **Markov matrix** (or **stochastic matrix**) is a square matrix M whose columns are probability vectors.

Definition

A **Markov chain** is a sequence of probability vectors $\vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$ such that $\vec{x}_{k+1} = M\vec{x}_k$ for some Markov matrix M.

Note: a Markov chain is determined by two pieces of information.

Steady-state vectors

- Given a Markov matrix M, does there exist a **steady-state** vector?

Steady-state vectors

- Given a Markov matrix M, does there exist a **steady-state** vector?
- This would be a probability vector $\vec{x}$ such that $M\vec{x} = \vec{x}$.

Steady-state vectors

- Given a Markov matrix M, does there exist a **steady-state** vector?
- This would be a probability vector $\vec{x}$ such that $M\vec{x} = \vec{x}$.
- Solve for steady-state in city-suburb example.

Voter preferences

- Suppose voter preferences (for parties "D", "R" and "L") shift around randomly via the Markov matrix

$$A = \begin{bmatrix} 0.70 & 0.10 & 0.30 \\ 0.20 & 0.80 & 0.30 \\ 0.10 & 0.10 & 0.40 \end{bmatrix}.$$

Voter preferences

- Suppose voter preferences (for parties "D", "R" and "L") shift around randomly via the Markov matrix

$$A = \begin{bmatrix} 0.70 & 0.10 & 0.30 \\ 0.20 & 0.80 & 0.30 \\ 0.10 & 0.10 & 0.40 \end{bmatrix}.$$

- e.g. 20% of supporters of "D" transition to "R" each election cycle.

Voter preferences

- Suppose voter preferences (for parties "D", "R" and "L") shift around randomly via the Markov matrix

$$A = \begin{bmatrix} 0.70 & 0.10 & 0.30 \\ 0.20 & 0.80 & 0.30 \\ 0.10 & 0.10 & 0.40 \end{bmatrix}.$$

- e.g. 20% of supporters of "D" transition to "R" each election cycle.
- Find steady-state.

Questions

Questions:

- How do we know a steady-state vector exists?

Questions

Questions:

- How do we know a steady-state vector exists?
- Does a steady-state vector always have nonnegative entries?

Questions

Questions:

- How do we know a steady-state vector exists?
- Does a steady-state vector always have nonnegative entries?
- Is a steady-state vector unique? Can you ever guarantee it?

Questions

Questions:

- How do we know a steady-state vector exists?
- Does a steady-state vector always have nonnegative entries?
- Is a steady-state vector unique? Can you ever guarantee it?
- Does the Markov chain always settle down to a steady-state vector?

Existence

Theorem

If M is a Markov matrix, there exists a vector $\vec{x} \neq \vec{0}$ such that $M\vec{x} = \vec{x}$.

Existence

Theorem

If M is a Markov matrix, there exists a vector $\vec{x} \neq \vec{0}$ such that $M\vec{x} = \vec{x}$.

Proof:

Existence

Theorem

If M is a Markov matrix, there exists a vector $\vec{x} \neq \vec{0}$ such that $M\vec{x} = \vec{x}$.

Proof:

- We're done if $M - I$ has a nontrivial kernel

Existence

Theorem

If M is a Markov matrix, there exists a vector $\vec{x} \neq \vec{0}$ such that $M\vec{x} = \vec{x}$.

Proof:

- We're done if $M - I$ has a nontrivial kernel (i.e., $M - I$ is not invertible).

Existence

Theorem

If M is a Markov matrix, there exists a vector $\vec{x} \neq \vec{0}$ such that $M\vec{x} = \vec{x}$.

Proof:

- We're done if $M - I$ has a nontrivial kernel (i.e., $M - I$ is not invertible).
- so we're done if $(M - I)^T$ has a nontrivial kernel

Existence

Theorem

If M is a Markov matrix, there exists a vector $\vec{x} \neq \vec{0}$ such that $M\vec{x} = \vec{x}$.

Proof:

- We're done if $M - I$ has a nontrivial kernel (i.e., $M - I$ is not invertible).
- so we're done if $(M - I)^T$ has a nontrivial kernel (but $(M - I)^T = M^T - I$).

Existence

Theorem

If M is a Markov matrix, there exists a vector $\vec{x} \neq \vec{0}$ such that $M\vec{x} = \vec{x}$.

Proof:

- We're done if $M - I$ has a nontrivial kernel (i.e., $M - I$ is not invertible).
- so we're done if $(M - I)^T$ has a nontrivial kernel (but $(M - I)^T = M^T - I$).
- But I can find a vector in kernel of $M^T - I$.

Uniqueness

- Can there be more than one steady-state vector?

Uniqueness

- Can there be more than one steady-state vector?
- How about $M = I_n$?

Long-term behavior

- Must the system "settle down" to a steady-state?

Long-term behavior

- Must the system "settle down" to a steady-state?
- How about $M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$?

Main theorem

Perron–Frobenius Theorem (circa 1910)

If M is a Markov matrix *with all positive entries*, then M **has** a **unique** steady-state vector, $\vec{x}$.

Main theorem

Perron–Frobenius Theorem (circa 1910)

If M is a Markov matrix *with all positive entries*, then M **has** a **unique** steady-state vector, $\vec{x}$. If $\vec{x}_0$ is any initial state, then $\vec{x}_k = M^k \vec{x}_0$ converges to $\vec{x}$ as $k \to \infty$.

Google's PageRank

- Problem: Given n interlinked webpages, rank them in order of "importance".

Google's PageRank

- Problem: Given n interlinked webpages, rank them in order of "importance".
- Assign the pages **importance scores** $x_1, x_2, \ldots, x_n \geq 0$.

Google's PageRank

- Problem: Given n interlinked webpages, rank them in order of "importance".
- Assign the pages **importance scores** $x_1, x_2, \ldots, x_n \geq 0$.
- Key insight: use the existing *link structure of the web* to determine importance.

Google's PageRank

- Problem: Given n interlinked webpages, rank them in order of "importance".
- Assign the pages **importance scores** $x_1, x_2, \dots, x_n \geq 0$.
- Key insight: use the existing *link structure of the web* to determine importance. A link to a page is like a vote for its importance.

Google's PageRank

- Problem: Given n interlinked webpages, rank them in order of "importance".
- Assign the pages **importance scores** $x_1, x_2, \ldots, x_n \geq 0$.
- Key insight: use the existing *link structure of the web* to determine importance. A link to a page is like a vote for its importance.
- How does this help with web searches?

Google's PageRank

- Problem: Given n interlinked webpages, rank them in order of "importance".
- Assign the pages **importance scores** $x_1, x_2, \ldots, x_n \geq 0$.
- Key insight: use the existing *link structure of the web* to determine importance. A link to a page is like a vote for its importance.
- How does this help with web searches?
- Working example: $n = 4$. Page 1 links to 2, 3, and 4. Page 2 links to 3 and 4. Page 3 links to 1. Page 4 links to 1 and 3.

Idea 1

- First attempt: let x_k equal the number of links to page k.

Idea 1

- First attempt: let x_k equal the number of links to page k.
- Do this in example.

Idea 1

- First attempt: let x_k equal the number of links to page k.
- Do this in example.
- Criticism: a link from an "important" page (like Yahoo) should carry more weight than a link from some random blog!

Idea 2

- Second attempt: let x_k equal the sum of the importance scores of all pages linking to page k.

Idea 2

- Second attempt: let x_k equal the sum of the importance scores of all pages linking to page k.
- Do this in example;

Idea 2

- Second attempt: let x_k equal the sum of the importance scores of all pages linking to page k.
- Do this in example; obtain a linear system.

Idea 2

- Second attempt: let x_k equal the sum of the importance scores of all pages linking to page k.
- Do this in example; obtain a linear system.
- Criticism 1: a webpage gets more "votes" (exerts more influence) if it has many outgoing links.

Idea 2

- Second attempt: let x_k equal the sum of the importance scores of all pages linking to page k.
- Do this in example; obtain a linear system.
- Criticism 1: a webpage gets more "votes" (exerts more influence) if it has many outgoing links.
- Criticism 2: this system only has the trivial solution!

Idea 3

- Third attempt (Brin and Page, late 90s): let x_k equal the sum of x_j/n_j, where

Idea 3

- Third attempt (Brin and Page, late 90s): let x_k equal the sum of x_j/n_j, where
- the sum is taken over all the pages j that link to page k, and

Idea 3

- Third attempt (Brin and Page, late 90s): let x_k equal the sum of x_j/n_j, where
- the sum is taken over all the pages j that link to page k, and
- n_j is the number of outoing links on page j.

Idea 3

- Third attempt (Brin and Page, late 90s): let x_k equal the sum of x_j/n_j, where
- the sum is taken over all the pages j that link to page k, and
- n_j is the number of outoing links on page j.
- That is, a page's number of votes is its importance score, and it gets split evenly among the pages it links to.

Idea 3

- Third attempt (Brin and Page, late 90s): let x_k equal the sum of x_j/n_j, where
- the sum is taken over all the pages j that link to page k, and
- n_j is the number of outoing links on page j.
- That is, a page's number of votes is its importance score, and it gets split evenly among the pages it links to.
- Do example.

Idea 3

- Third attempt (Brin and Page, late 90s): let x_k equal the sum of x_j/n_j, where
- the sum is taken over all the pages j that link to page k, and
- n_j is the number of outoing links on page j.
- That is, a page's number of votes is its importance score, and it gets split evenly among the pages it links to.
- Do example.
- Write in matrix form.

Idea 3

- Third attempt (Brin and Page, late 90s): let x_k equal the sum of x_j/n_j, where
- the sum is taken over all the pages j that link to page k, and
- n_j is the number of outoing links on page j.
- That is, a page's number of votes is its importance score, and it gets split evenly among the pages it links to.
- Do example.
- Write in matrix form.
- Solve, and rank the pages by importance.

Summary

- Summary: given a web with n pages, construct an $n \times n$ matrix A as:

$$a_{ij} = \begin{cases} 1/n_j, & \text{if page } j \text{ links to page } i \\ 0, & \text{otherwise} \end{cases},$$

where n_j is the number of outgoing links on page j.

Summary

- Summary: given a web with n pages, construct an $n \times n$ matrix A as:

$$a_{ij} = \begin{cases} 1/n_j, & \text{if page } j \text{ links to page } i \\ 0, & \text{otherwise} \end{cases},$$

where n_j is the number of outgoing links on page j.

- Sum of jth column is $n_j/n_j = 1$, so A is a Markov matrix.

Summary

- Summary: given a web with n pages, construct an $n \times n$ matrix A as:

$$a_{ij} = \begin{cases} 1/n_j, & \text{if page } j \text{ links to page } i \\ 0, & \text{otherwise} \end{cases},$$

  where n_j is the number of outgoing links on page j.
- Sum of jth column is $n_j/n_j = 1$, so A is a Markov matrix.
- The ranking vector $\vec{x}$ solves $A\vec{x} = \vec{x}$.

Summary

- Summary: given a web with n pages, construct an $n \times n$ matrix A as:

$$a_{ij} = \begin{cases} 1/n_j, & \text{if page } j \text{ links to page } i \\ 0, & \text{otherwise} \end{cases},$$

  where n_j is the number of outgoing links on page j.
- Sum of jth column is $n_j/n_j = 1$, so A is a Markov matrix.
- The ranking vector $\vec{x}$ solves $A\vec{x} = \vec{x}$.
- Possible issues:

Summary

- Summary: given a web with n pages, construct an $n \times n$ matrix A as:

$$a_{ij} = \begin{cases} 1/n_j, & \text{if page } j \text{ links to page } i \\ 0, & \text{otherwise} \end{cases},$$

  where n_j is the number of outgoing links on page j.
- Sum of jth column is $n_j/n_j = 1$, so A is a Markov matrix.
- The ranking vector $\vec{x}$ solves $A\vec{x} = \vec{x}$.
- Possible issues:
- Existence of solution with nonnegative entries?

Summary

- Summary: given a web with n pages, construct an $n \times n$ matrix A as:

$$a_{ij} = \begin{cases} 1/n_j, & \text{if page } j \text{ links to page } i \\ 0, & \text{otherwise} \end{cases},$$

  where n_j is the number of outgoing links on page j.
- Sum of jth column is $n_j/n_j = 1$, so A is a Markov matrix.
- The ranking vector $\vec{x}$ solves $A\vec{x} = \vec{x}$.
- Possible issues:
- Existence of solution with nonnegative entries?
- Non-unique solutions?

Google's PageRank

- PF Theorem guarantees a unique steady-state vector if the entries of A are strictly positive.

Google's PageRank

- PF Theorem guarantees a unique steady-state vector if the entries of A are strictly positive.
- Brin–Page: replace A with

$$B = 0.85A + 0.15(\text{matrix with every entry } 1/n).$$

Google's PageRank

- PF Theorem guarantees a unique steady-state vector if the entries of A are strictly positive.
- Brin–Page: replace A with

$$B = 0.85A + 0.15(\text{matrix with every entry } 1/n).$$

- B is a Markov matrix.

Google's PageRank

- PF Theorem guarantees a unique steady-state vector if the entries of A are strictly positive.
- Brin–Page: replace A with

$$B = 0.85A + 0.15(\text{matrix with every entry } 1/n).$$

- B is a Markov matrix.
- PF Theorem says B has a unique steady-state vector, $\vec{x}$.

Google's PageRank

- PF Theorem guarantees a unique steady-state vector if the entries of A are strictly positive.
- Brin–Page: replace A with

$$B = 0.85A + 0.15(\text{matrix with every entry } 1/n).$$

- B is a Markov matrix.
- PF Theorem says B has a unique steady-state vector, $\vec{x}$.
- So $\vec{x}$ can be used for rankings!

Stochastic interpretation of PageRank

- What does this have to do with Markov chains?

Stochastic interpretation of PageRank

- What does this have to do with Markov chains?
- Brin and Page considered web surfing as a stochastic process:

Stochastic interpretation of PageRank

- What does this have to do with Markov chains?
- Brin and Page considered web surfing as a stochastic process:

Quote

PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page.

Stochastic interpretation of PageRank

- What does this have to do with Markov chains?
- Brin and Page considered web surfing as a stochastic process:

Quote

PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page.

- i.e., surfer clicks on a link on the current page with probability 0.85; opens up a random page with probability 0.15.

Stochastic interpretation of PageRank

- What does this have to do with Markov chains?
- Brin and Page considered web surfing as a stochastic process:

Quote

PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page.

- i.e., surfer clicks on a link on the current page with probability 0.85; opens up a random page with probability 0.15.
- A page's rank is the probability the random user will end up on that page, OR, equivalently

Stochastic interpretation of PageRank

- What does this have to do with Markov chains?
- Brin and Page considered web surfing as a stochastic process:

Quote

PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page.

- i.e., surfer clicks on a link on the current page with probability 0.85; opens up a random page with probability 0.15.
- A page's rank is the probability the random user will end up on that page, OR, equivalently
- the fraction of time the random user spends on that page in the long run.