eigenvalues, markov matrices, and the power method

Slides by Olson. Some taken loosely from Jeff Jauregui, Some from Semeraro

L. Olson

Department of Computer Science
University of Illinois at Urbana-Champaign

objectives

- Create a stochastic matrix (or Markov matrix) that represents the probability of moving from one state to the next
- Establish properties of the Markov Matrix
- Find the steady state of a stochastic matrix
- Relate the steady state to an eigenvecture
- Find important eigenvectors with the Power Method

random transitions

- Given a system of “states“, we want to model the transition from state to state over time.
- Let n be the number of states
- So at time k the system is represented by $x_k \in \mathbb{R}^n$.
- $x_k^{(i)}$ is the probability of being in state i at time k

Definition

A probability vector is a vector of positive entries that sum to 1.0.

Definition

A Markov matrix is a square matrix M with columns that are probability vectors. So the entries of M are positive and the column sums are 1.0.

Definition

A Markov Chain is a sequence of probability vectors $x_0, x_1, \ldots, x_k, \ldots$ such that

$$x_{k+1} = Mx_k$$

for some Markov Matrix M

markov chains

- Does a steady-state exist?
- Does a steady state depend on the initial state?
- Will x_{k+1} be a probability vector if x_k is a probability vector?
- Is the steady state unique?

Theorem

Let M be a Markov Matrix. Then there is a vector $x \neq 0$ such that $Mx = x$.

Proof?

- M^T is singular. Why?
- So there is an x such that $M^T x = x$
- or so that $(M^T - I)x = 0$
- Thus $M - I$ is singular. Why?

goal

- Find $x = Ax$ and the elements of x are the probability vector (Basketball Ranking, Google Page Rank, etc).

power method

Suppose that A is $n \times n$ and that the eigenvalues are ordered:

$$|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \cdots \geqslant |\lambda_n|$$

Assuming A is nonsingular, we have a linearly independent set of v_i such that $Av_i = \lambda_i v_i$.

Goal

Computing the value of the largest (in magnitude) eigenvalue, λ_1.

power method

Take a guess at the associated eigenvector, x_0. We know

$$x^{(0)} = c_1 v_1 + \cdots + c_n v_n$$

Since the guess was random, start with all $c_j = 1$:

$$x^{(0)} = v_1 + \cdots + v_n$$

Then compute

$$\begin{aligned} x^{(1)} &= Ax^{(0)} \\ x^{(2)} &= Ax^{(1)} \\ x^{(3)} &= Ax^{(2)} \\ &\vdots \\ x^{(k+1)} &= Ax^{(k)} \end{aligned}$$

power method

Or $x^{(k)} = A^k x^{(0)}$. Or

$$\begin{aligned} x^{(k)} &= A^k x^{(0)} \\ &= A^k v_1 + \cdots + A^k v_n \\ &= \lambda_1^k v_1 + \ldots \lambda_n^k v_n \end{aligned}$$

And this can be written as

$$x^{(k)} = \lambda_1^k \left(v_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n \right)$$

So as $k \to \infty$, we are left with

$$x^{(k)} \to \lambda^k v_1$$

the power method (with normalization)

```
for k = 1 to kmax
    y = Ax
    r = φ(y)/φ(x)
    x = y/‖y‖∞
```

- often $\phi(x) = x_1$ is sufficient
- r is an estimate of the eigenvalue; x the eigenvector

inverse power method

- We now want to find the smallest eigenvalue
- $Av = \lambda v \quad \Rightarrow \quad A^{-1}v = \frac{1}{\lambda}v$
- So “apply” power method to A^{-1} (assuming a distinct smallest eigenvalue)
- $x^{(k+1)} = A^{-1}x^{(k)}$
- Easier with $A = LU$
- Update RHS and backsolve with U:

$$Ux^{(k+1)} = L^{-1}x^{(k)}$$

theory

Theorem

Perron-Frobenius If M is a Markov matrix with positive entries, then M has a unique steady-state vector x.

Theorem

Perron-Frobenius Corollary Given an initial state x_0, *then* $x_k = M^k x_0$ *converges to* x.

pagerank

Example

Problem: Consider n linked webpages. Rank them.

- Let $x_1, \ldots, x_n \geqslant 0$ represent *importance*
- A link to a page increases the perceived importance of a webpage

Example

Try $n = 4$.

- page 1: 2,3,4
- page 2: 3,4
- page 3: 1
- page 4: 1,3

page rank

First attempt

- Let x_k be the number of links to page k
- Problem: a link from an important page like The NY Times has no more weight than `lukeo.cs.illinois.edu`

page rank

Second attempt

- Let x_k be the sum of importance scores of all pages that link to page k
- Problem: a webpage has more influence simply by having more outgoing links
- Problem: the linear system is trivial (oops!)

page rank

Third attempt (Brin/Page '90s)

- Let n_j be the number of outgoing links on page j
- Let

$$x_k = \sum_{j \text{ linking to } k} \frac{x_j}{n_j}$$

- The influence of a page is its importance. It is split evenly to the pages it links to.

Example

Let A be an $n \times n$ matrix as

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

page rank

- Sum of column j is $n_j/n_j = 1$, so A is a Markov Matrix
- Problem: does not guarantee a unique x s.t. $Ax = x$
- Brin-Page: Use instead

$$A \leftarrow 0.85A + 0.15$$

- Still a Markov Matrix
- Now has all positive entries
- Guarantees a unique solution

page rank

$$A \leftarrow 0.85A + 0.15$$

- What does this mean though?
- This defines a stochastic process: “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 *bakc*, but eventually gets bored and starts on another random page.”
- So a surfer clicks on a link on the current page with probability 0.85 and opens a random page with probability 0.15.
- PageRank is the probability that the random user will end up on that page