Announcements

Tuesday, February 27

Written Assignment: *Prioritize quality over quantity.*

Setting: Representing a star-up, you are writing a report to a big company to offer your services. **In 2-3 pages, cover the following issues:**

- Give a 1/2 page summary of the proposal targeted to a layman
- What is the problem of the company (be creative)
- What is your proposed solution
- *Why does your solution works* (technical background)
- How will you implement the solution

Suggested companies/projects:

- Pixar's 3D simulations, Urban Planning, Tracing the trayectory of Falcon 9

Grading Scheme:

1. Correctness: 4pts
2. Clarity: 4pts
3. Creativity: 3pts

Deadline: Emailed in PDF by April 20th

Group Sizes: 3-4 persons

- Suggested sources: Sections
- 2.6,2.7,3.3,4.8,4.9,5.6,5.7,5.8, 6.4,6.6,6.8,7.5,10.1,10.3,10.4
- *Not exclusively* Chapter 1

Application

Stochastic Matrices and PageRank

Stochastic Matrices

These arise very commonly in modeling of probabalistic phenomena (Markov chains), where they are also called **transition matrices**.

Some examples:

- Matrices from the population dynamics
- Matrices from the equilibrium-prices economies

Definition

A square matrix A is **stochastic** if all of its entries are nonnegative, and the sum of the entries of each column is 1.

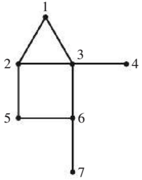
We say A is **regular** if, for some k, all entries of A^k are positive.

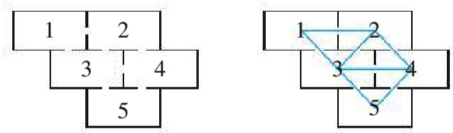
Definition

A **steady-state vector** v of A is a non-zero vector with *entries summing to 1* and such that $Av = v$.

Random walks on graphs (a.k.a Mouse on a maze)

A mouse moves freely between rooms/states = selects any with equal probability.





$$P = \begin{array}{cc} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} & \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/4 & 0 & 1/2 & 0 & 0 \\ 1/2 & 1/3 & 0 & 1 & 0 & 1/3 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/4 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} \end{array}$$

$$P = \begin{array}{cc} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} & \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/3 & 0 & 1/3 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/4 & 1/3 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \end{array}$$

- *Initial state:* The mouse is located at some room i: probabilities $v_0 = (x_1, \vdots, x_5)$.
- Probability mouse starts at room 1 is x_1 item *Transition matrix:* $v_{n+1} = Av_n$ means that A dictates how probabilities change.
- Probability mouse is at room 3 after n steps of the walk: third entry of v_n.

Non-regular transition matrix

Disconnected states

Consider the following 'transition graph':

The transition matrix is $\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$.

Both $\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, are eigenvectors with eigenvalue 1.

So there is more than one steady-state vector!

Red Box has kiosks all over where you can rent movies. You can return them to any other kiosk.

- *ij entry of A:* probability that a movie rented from location j is returned to location i.

For example, if there are three locations, maybe

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

30% probability a movie rented from location 3 gets returned to location 2

On day n: x_n, y_n, z_n are the numbers of movies in locations $1, 2, 3$, respectively, and $v_n = (x_n, y_n, z_n)$.

Probabilistic Intuition

If at opening day the movies are distributed according to v_0 then, **on average**:

$$v_n = Av_{n-1} = A^2 v_{n-2} = \cdots = A^n v_0.$$

Eigenvalues of Stochastic Matrices

Fact: 1 is an eigenvalue of a stochastic matrix.

Why? If A is stochastic, then 1 is an eigenvalue of A^T:

$$\begin{pmatrix} .3 & .3 & .4 \\ .4 & .4 & .2 \\ .5 & .3 & .2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Lemma

A and A^T have the same eigenvalues.

Proof: $\det(A - \lambda I) = \det\big((A - \lambda I)^T\big) = \det(A^T - \lambda I)$, so they have the same characteristic polynomial.

Note: This doesn't give a new procedure for finding an eigenvector with eigenvalue 1; it only shows one exists.

Stonger fact: if $\lambda \neq 1$ is an eigenvalue of a *regular* stochastic matrix, then $|\lambda| < 1$.

Eigenvalues of Stochastic Matrices
Continued

So: If λ is an eigenvalue of A then it is an eigenvalue of A^T.

$$\text{eigenvector } v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \lambda v = A^T v \implies \lambda x_j = \underbrace{\boxed{\sum_{i=1}^n a_{ij} x_i.}}_{j\text{th entry of } A^T v}$$

Choose x_j with the largest absolute value, so $|x_i| \le |x_j|$ for all i.

$$|\lambda| \cdot |x_j| = \left| \sum_{i=1}^n a_{ij} x_i \right| \le \sum_{i=1}^n \overbrace{a_{ij}}^{\text{positive}} \cdot |x_i| \le \sum_{i=1}^n a_{ij} \cdot \underbrace{|x_j|}_{\ge |x_i|} = \overbrace{1}^{= \sum_i a_{ij}} \cdot |x_j|,$$

so $|\lambda| \le 1$.

Conclusion: if $\lambda \neq 1$ is an eigenvalue of a stochastic matrix with all entries positive, then $|\lambda| < 1$. This proof is adapted for *regular* matrices.

Dynamical systems

Except for the 1-eigenspace, all the others are shriking!
What happens in the long run?

Diagonalizable Stochastic Matrices

Example from §5.3

Let $A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$. This is a regular stochastic matrix.

We saw last time that A is diagonalizable:

$$A = PDP^{-1} \quad \text{for} \quad P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Change of basis: Let $w_1 = \binom{1}{1}$ and $w_2 = \binom{1}{-1}$ be the columns of P.

$$A^n = c_1 w_1 + \frac{c_2}{2^n} w_2.$$

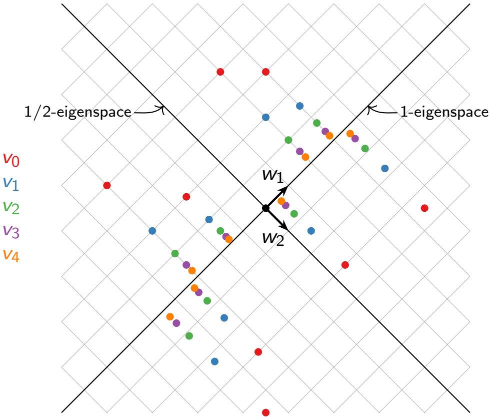
When n is large, the second term disappears, so $A^n x$ approaches $c_1 w_1$, which is an *eigenvector with eigenvalue* 1 (assuming $c_1 \neq 0$).

So all vectors get "sucked into the 1-eigenspace," which is spanned by $w_1 = \binom{1}{1}$.

Diagonalizable Stochastic Matrices

Example, continued

Recall: $A^n = PD^nP^{-1}$ acts on the usual coordinates of v_0 in the same way that D^n acts on the $\mathcal{B}$-coordinates, where $\mathcal{B} = \{w_1, w_2\}$.



All vectors get "sucked into the 1-eigenspace."

Diagonalizable Stochastic Matrices

The Red Box matrix $A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$ is diagonalizable

$$A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & .1 & 0 \\ 0 & 0 & -.2 \end{pmatrix} P^{-1} = PDP^{-1}.$$

Hence it is easy to compute the powers of A:

$$A^n = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & (.1)^n & 0 \\ 0 & 0 & (-.2)^n \end{pmatrix} P^{-1} = PD^n P^{-1}.$$

Let w_1, w_2, w_3 be the columns of P, i.e. the eigenvectors of P with respective eigenvalues $1, .1, -.2$. Let $\mathcal{B} = \{w_1, w_2, w_3\}$.

If w_1, w_2, w_3 are the column vectors of P then

$$A^n x = c_1 w_1 + (.1)^n c_2 w_2 + (-.2)^n c_3 w_3 \to c_1 w_1$$

If A is the Red Box matrix, and v_n is the vector representing the number of movies in the three locations on day n, then

$$v_{n+1} = Av_n.$$

For any starting distribution v_0 of videos in red boxes, after enough days, the distribution v ($= v_n$ for n large) is an eigenvector with eigenvalue 1:

$$Av = v.$$

In other words, eventually the number of movies in each kiosk doesn't change much.

Moreover, we know exactly what v is: a multiple of w_1

- The entries in v have to sum up to the number of intial movies (same sum as entries in v_0.

(Remember the total number of videos never changes.) Presumably, Red Box

really does have to do this kind of analysis to determine how many videos to put in each box.

Find the actual Steady State w_1

Red Box example

If one computes $\mathrm{Nul}(A - I)$ and find that $w' = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}$

is an eigenvector with eigenvalue 1.

Then, to get a steady state, divide by $18 = 7 + 6 + 5$ to get

$$w = \frac{1}{18} \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix} \sim (0.39, 0.33, 0.28).$$

The long-run

So if you start with 100 total movies, eventually you'll have $100w = (39, 33, 28)$ movies in the three locations, every day.

The time spent on a state

Regardless of the intital location of a particular movie. Eventually, that movie will get 'returned' 39% of the times at location 1, 33% at location 2, and 28% at location 3.

Perron–Frobenius Theorem

These conclusions apply to *any* regular stochastic matrix—whether or not it is diagonalizable!

Perron–Frobenius Theorem

If A is a regular stochastic matrix, then it admits a unique steady state vector w, which spans the 1-eigenspace.

Moreover, for any vector v_0 with entries summing to some number c, the iterates $v_1 = Av_0$, $v_2 = Av_1$, ..., $v_n = Av_{n-1}$, ..., approach cw as n gets large.

Translation:

- The *1-eigenspace* of a regular stochastic matrix A *is a line*.
- The *vector w* has entries that sum to 1, and are *strictly positive*!
- Eventually, the *movies arrange* themselves according to the *steady state percentage*, i.e., $v_n \to cw$.

  (The sum c of the entries of v_0 is the total number of movies)

Google's PageRank

Internet searching in the 90's was a pain. Yahoo or AltaVista would scan pages for your search text, and just list the results with the most occurrences of those words.

Not surprisingly, the more unsavory websites soon learned that by putting popular words a million times in their pages, they could show up first on popular searches.

Larry Page and Sergey Brin invented a way to rank pages by *importance*. They founded Google based on their algorithm.

Here's how it works. (roughly)

Reference:

http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

Google's PageRank: The Importance Rule

Each webpage has an associated importance, or **rank**. This is a positive number.

The Importance Rule

If page P *links* to n other pages $Q_1, Q_2, \ldots, Q_n$, then each Q_i *should inherit* $\frac{1}{n}$ of P's *importance*.

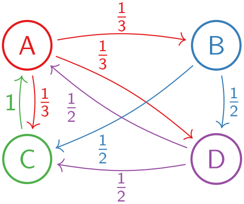
- A very important page links to your webpage: then your webpage is important.
- A ton of unimportant pages link to your webpage: then it's still important.
- But if only one crappy site links to yours, your page isn't important.

Random surfer interpretation

A "random surfer" just randomly clicks on link after link. The pages she *spends the most time* on should be *the most important*. **Stochastic terms:** random walk on the graph of hiperlinks. Look for steady-state vector!

The Importance Matrix

Consider the following Internet with only four pages. Links are indicated by arrows.



In terms of matrices, if $v = (a, b, c, d)$ is the vector containing the ranks a, b, c, d of the pages A, B, C, D, then

importance matrix: *ij* entry is importance page *j* passes to page *i*

$$\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} c + \frac{1}{2}d \\ \frac{1}{3}a \\ \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d \\ \frac{1}{3}a + \frac{1}{2}b \end{pmatrix} \overset{\text{Importance Rule}}{=} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

The 25 Billion Dollar Eigenvector

Observations:

- The importance matrix is a stochastic matrix!
- The rank vector is an eigenvector with eigenvalue 1

Random surfer interpretation: If a random surfer has probability (a, b, c, d) to be on page A, B, C, D, respectively, then after clicking on a random link, the probability he'll be on each page is

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

The rank vector is a *steady state* for the importance matrix : it's the probability vector (a, b, c, d) such that, after clicking on a random link, the random surfer will have the *same probability* of being on each page.

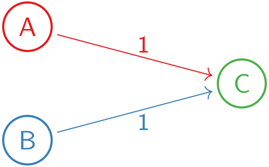
So, the important (high-ranked) pages are those where a random surfer will end up most often.

Problems with the Importance Matrix

Dangling pages

Observation: the importance matrix is *not* regular: it's only nonnegative. So we can't apply the Perron–Frobenius theorem. *How does this cause problems?*

Consider the following Internet:



The importance matrix is $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$. is not stochastic!

$$f(\lambda) = \det \begin{pmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 1 & 1 & -\lambda \end{pmatrix} = -\lambda^3.$$

and 1 is not even an eigenvalue: there is no rank vector!

The Google Matrix (Page and Brin's solution)

Fix p in $(0, 1)$, called the **damping factor**. (A typical value is $p = 0.15$.)

The **Google Matrix** is

$$M = (1 - p) \cdot A + p \cdot B \quad \text{where} \quad B = \frac{1}{N} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

N is the total number of pages, and A is the importance matrix.

- Random surfer interpretation: with probability p the surfer gets bored and starts over on a completely random page.

Fact

The PageRank vector is the steady state for the Google Matrix.

This exists and has positive entries by the Perron–Frobenius theorem.
The hard part is calculating it: the Google matrix has 1 gazillion rows.