Lesson 7

Google page rank

Theorem 3.2.4. *(***Perron's theorem for positive matrices***) If* $A > 0$ *then* A *has a* dominant eigenvalue λ_A *with the following properties:*

1. $\lambda_A > 0$ *and it has an eigenvector* $\boldsymbol{u}_A > 0$ *(has all positive entries) called a* dominant eigenvector *of* A.
2. $\mathrm{AM}(\lambda_A) = 1$.
3. *For any other eigenvalue* μ *of* A*, all its eigenvectors have at least one negative coordinate, and* $|\mu| < \lambda_A$.

Theorem 3.2.5. (**Perron's theorem for positive Markov matrices**) *If A is positive Markov then A has all the properties stated in Theorem 3.2.4, and in addition:*

1. *$\lambda_A = 1$ and thus if μ is another eigenvalue of A, then $|\mu| < 1$.*
2. *If $\boldsymbol{w}_0 \geq 0$ and $\boldsymbol{w}_k = A^k \boldsymbol{w}_0$, then $\lim_{k\to\infty} A^k \boldsymbol{w}_0 = c\boldsymbol{u}_A$ where $c \geq 0$. In other words, the in the long run, $\boldsymbol{w}_k$ tends to a non-negative multiple of $\boldsymbol{u}_A$.*

[2]

3) if w_0 and u_A are probability vectors, $c=1$

Proof using Perron's th for positive matrices

1) The eigenvalues of A and A^T are the same

A^T is positive and we know $A^T \mathbb{1} = \mathbb{1}$

so $\lambda = 1$ must be the dominant eigenvalue of A^T (why? Because only the dominant eigenvalue of A^T can have a positive eigen vector, by Perron's theorem for positive matrices)

The eigenvalues of A^T and A are the same (see hw 1) so $\lambda = 1$ is the dominant eigenvalue of A.

2) Proof only in the case A is diagonalizable:

$$w_k = A^k w_0 = P D^k P^{-1} w_0 = P D^k \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = P \begin{bmatrix} c_1 \lambda_1^k & 0 & \cdots & 0 \\ 0 & c_2\lambda_2^k & & \\ \vdots & & \ddots & \\ 0 & & & c_n\lambda_n^k \end{bmatrix}$$

$$= c_1 \lambda_1^k u_1 + c_2 \lambda_2^k u_2 + \cdots + c_n \lambda_n^k$$

$\lambda_1 = 1$

$|\lambda_i| < 1 \quad i > 1$

$$= c_1 u_1 + \text{stuff that goes to } 0$$

u_1 is dominant eigenvector i.e eigenvector for $\lambda = 1$

what can we say about c_1?

$$\lim_{k \to +\infty} A^k w_0 = c_1 v_1 \quad \text{(from previous page)}$$

$$* \quad \lim_{k \to +\infty} w_0^T (A^k)^T = c_1 v_1^T \quad \text{(Taking transposes)}$$

$$1^T A^k = 1^T \quad \text{(because } A^k \text{ is Markov)}$$

$$(A^k)^T 1 = 1 \quad \text{(Taking transposes)}$$

$$w_0^T (A^k)^T 1 = w_0^T 1 \quad \text{(multiplying by } w_0^T\text{)}$$

$$\lim_{k \to +\infty} w_0^T (A^k)^T 1 = \lim_{k \to +\infty} w_0^T 1 \quad \text{(Taking limits)}$$

$$c_1 v_1^T 1 = w_0^T 1 \quad \text{(by } *\text{)}$$

If v_1 and w_0 are probability vectors

$$v_1^T \cdot 1 = [x_1 \cdots x_n] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = x_1 + \cdots + x_n = 1$$

$$w_0^T 1 = [y_1 \cdots y_n] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = y_1 + y_2 + \cdots + y_n = 1$$

Therefore $c_1 \cdot 1 = 1$

What if A is not positive?

Ex $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is Markov

eigenvalues $\lambda = 1, 1$ no dominant eigenvalue

$M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is also Markov

$$M^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

what is $\lim\limits_{n \to +\infty} M^n \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix} = ?$

Def: A is regular Markov if A is Markov and A^k is positive for some k

Th: if A is regular Markov th 3.2.5 is still true for A

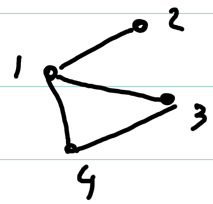
A^k is positive Markov, the eigenvalues of A are 1 (because A is Markov) $\lambda_2 \dots \lambda_n$

the eigenvalues of A^k are $1^k\ \lambda_2^k \dots \lambda_n^k$

since A^k is positive Markov we know

$|\lambda_i^k| < 1$ so $|\lambda_i|^k < 1$ so $|\lambda_i| < 1$

Adjacency matrix of a Graph



Vertices 1, 2, 3, 4

Edges {1,2} {1,3} {1,4} {3,4}

How can we represent this information using a matrix ?

$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & & & & \\ 2 & & & & \\ 3 & & & & \\ 4 & & & & \end{array} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

$$a_{ij} = \begin{cases} 0 & \text{if } \{i\ j\} \text{ not an edge} \\ 1 & \text{if } \{i\ j\} \text{ is an edge} \end{cases}$$

non negative symmetric

Talk about hw 2 #4 :

If G has n vertices, how many subgraphs does it have with 3 vertices ? $\binom{n}{3} = \dfrac{n!}{3!(n-3)!}$ $O(n^3)$

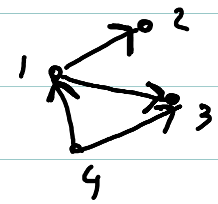
Look at all of them and decide if any is a triangle takes $O(n^3)$

Adjacency matrix is $n \times n$

Computing A^2 takes ? OPEN PROBLEM IN CS

but we can do better than n^3

Adjacency matrix of a directed graph



Vertices 1, 2, 3, 4

Edges (1 2) (1 3) (4 3) (4, 1)

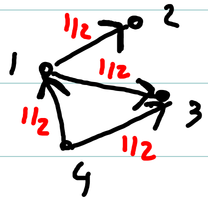
12

How can we represent this information using a matrix?

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$a_{ij} = \begin{cases} 1 & \text{if } (j\ i) \text{ is an edge} \\ 0 & \text{ow} \end{cases}$$

Adjacency matrix of a weighted directed graph



Vertices 1, 2, 3, 4

Edges (1 2) (1 3) (4 3) (4, 1)

If a vertex has k outgoing edges each is given weight $\frac{1}{k}$

$$
\begin{array}{c c}
 & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{array}
$$

P_{LS}

NOT POSITIVE

NOT MARKOV

Google page rank algorithm:

Goal: WWW pages relevant for a search form a weighted directed graph with n nodes (NODES = PAGES EDGES = LINKS) "represented" by some positive Markov Matrix $M = (p_{ij})$

STATES = pages p_{ij}: probability that if I am on page j at time t_k I will be at page i at time t_{k+1}

How do I go from page j to i?

1) Follow a link, each with equal probability

Problem, in general (p_{ij}) not positive not Markov (if there are vertices with no outgoing edges).

2) With probability $\frac{5}{6}$ follow a link (all equally probable); with probability $\frac{1}{6}$ jump to another page (all equally probable)

If there are no links jump with probability 1

Suppose that I can start on any page with equal probability so $w_0 = \begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$

In the long run which page(s) am I more likely to end on?

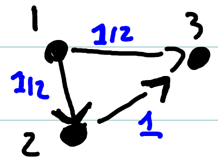
Answer $\lim_{k \to +\infty} M^k w_0 = U_1$ page rank vector

U_1 = probability vector for eigenvalue $\lambda = 1$ of M

Pages I am more likely to end up on are more important and should be displayed first.

Example:

suppose I have identified pages relevant to my search and their graph looks like this. Rules to build H



on page/vertex J : with probability $\frac{5}{6}$ follow a link (all equally probable); with probability $\frac{1}{6}$ jump to another page (all equally probable)

If there are no links jump with probability 1

3 states : S_1 : on page 1 S_2: on page 2

S_3 = on page 3

starting at

$$\text{going to}\quad \begin{array}{c c} & \begin{array}{ccc} \text{pg 1} & \text{pg 2} & \text{pg 3} \end{array} \\ \begin{array}{c} \text{pg 1} \\ \text{pg 2} \\ \text{pg 3} \end{array} & \begin{bmatrix} \frac{1}{6}\cdot\frac{1}{3} & \frac{1}{6}\cdot\frac{1}{3} & \frac{1}{3} \\ \frac{5}{6}\cdot\frac{1}{2}+\frac{1}{6}\cdot\frac{1}{3} & \frac{1}{6}\cdot\frac{1}{3} & \frac{1}{3} \\ \frac{5}{6}\cdot\frac{1}{2}+\frac{1}{6}\cdot\frac{1}{3} & \frac{5}{6}+\frac{1}{6}\cdot\frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{array}$$

Check it adds to 1

This is positive Markov

$$M = \begin{bmatrix} 2/36 & 1/18 & 1/3 \\ 17/36 & 1/18 & 1/3 \\ 17/36 & 16/18 & 1/3 \end{bmatrix}$$

if I start at pages 1, 2, 3 with equal probability and move to the next page and next page according to the rules in the previous slide what is going to happen in the long run?

want $\lim_{k \to +\infty} M^k \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = U_1$

U_1 = probability eigenvector for $\lambda = 1$

```
julia> using LinearAlgebra

julia> M=[2/36 1/18 1/3; 17/36 1/18 1/3; 17/36 16/18 1/3]
3x3 Matrix{Float64}:
 0.0555556  0.0555556  0.333333
 0.472222   0.0555556  0.333333
 0.472222   0.888889   0.333333

julia> eigenvals(M)
ERROR: UndefVarError: eigenvals not defined
Stacktrace:
 [1] top-level scope
   @ REPL[3]:1

julia> eigvals(M)
3-element Vector{ComplexF64}:
 -0.2777777777777778 - 0.19641855032959663im
 -0.2777777777777778 + 0.19641855032959663im
  0.9999999999999989 + 0.0im

julia> eigvecs(M)
3x3 Matrix{ComplexF64}:
 -0.471405+0.333333im  -0.471405-0.333333im  0.320215+0.0im
 -0.235702-0.333333im  -0.235702+0.333333im  0.453638+0.0im
  0.707107-0.0im        0.707107+0.0im        0.83167+0.0im

julia> 0.32/1.606
0.199252801992528

julia> 0.454/1.606
0.28268991282689915

julia> 0.832/1.606
0.5180572851805728

julia>
```

✓ 1

v_1

not a probability vector

$0.32 + 0.45 + 0.83 = 1.6$

$$v_1 = \begin{pmatrix} \frac{0.32}{1.6} \\ \frac{0.45}{1.6} \\ \frac{0.83}{1.6} \end{pmatrix} \approx \begin{pmatrix} 0.2 \\ 0.3 \\ 0.5 \end{pmatrix}$$

dominant eigenvalue $\lambda = 1$

dominant probability eigenvector $\begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$

$$\lim_{n \to +\infty} M^n \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$$

Def: the only dominant eigenvector that is a probability vector is called the page rank vector of the system of web pages.

Display page 3, then 2, then 1

Question: couldn't I just assume that a vertex / web page with more incoming edges is more important?

A link from an important page should boost the rank more than a link from a less important page