Session 6 bis: Markov Chains and PageRank

Optimization and Computational Linear Algebra for Data Science

Léo Miolane

Be accurate !

Let x, v be vectors, S a subspace of $\mathbb{R}^n$ and M an $n \times n$ matrix.

- $x = S$ or $x \subset S$
- $S \in \mathbb{R}^n$
- $\mathrm{Span}(x, v) = \{ax + bv\}$
- $\dim(M)$ or $\dim(x)$
- $\mathrm{Ker}(M) = 0$
- $x + M$

Contents

Markov chains

An example

Stochastic matrices

Definition

A matrix $P \in \mathbb{R}^{n \times n}$ is said to be *stochastic* if:

1. $P_{i,j} \geq 0$ for all $1 \leq i, j \leq n$.
2. $\sum_{i=1}^{n} P_{i,j} = 1$, for all $1 \leq j \leq n$.

Probability vectors

The key equation

Proposition

For all $t \geq 0$

$$x^{(t+1)} = Px^{(t)} \quad \text{and consequently,} \quad x^{(t)} = P^t x^{(0)}.$$

Long-term behavior

Perron-Frobenius Theorem

Invariant measure

Definition

A vector $\mu \in \Delta_n$ is called an invariant measure for the transition matrix P if

$$\mu = P\mu,$$

i.e. if μ is an eigenvector of P associated with the eigenvalue 1.

Perron-Frobenius Theorem

Theorem

Let P be a stochastic matrix such that there exists $k \geq 1$ such that all the entries of P^k are strictly positive. Then the following holds:

1. 1 is an eigenvalue of P and there exists an eigenvector $\mu \in \Delta_n$ associated to 1.
2. The eigenvalue 1 has multiplicity 1: $\mathrm{Ker}(P - \mathrm{Id}) = \mathrm{Span}(\mu)$.
3. For all $x \in \Delta_n$, $P^t x \xrightarrow[t \to \infty]{} \mu$.

Consequence

Corollary

Let P be a stochastic matrix such that there exists $k \geq 1$ such that all the entries of P^k are strictly positive.
Then there exists a unique invariant measure μ and for all initial condition $x^{(0)} \in \Delta_n$,

$$x^{(t)} = P^t x^{(0)} \xrightarrow[t \to \infty]{} \mu.$$

Proof: Geometrical observations

Proof: contraction

We will prove the theorem in the case where $P_{i,j} > 0$ for all i, j.

Lemma

The mapping

$$\begin{array}{rccc} \varphi : & \Delta_n & \to & \Delta_n \\ & x & \mapsto & Px \end{array}$$

is a contraction mapping for the ℓ_1-norm: there exists $c \in (0, 1)$ such that for all $x, y \in \Delta_n$:

$$\|Px - Py\|_1 \leq c\|x - y\|_1.$$

Geometric picture

Proof of Perron-Frobenius

Proof of Perron-Frobenius

PageRank

Ordering the Web

Naive attempt

First idea: rank the webpages according to their number of *incomming links*. (The more incomming links, the more the webpage is important).

The random surfer

PageRank Algorithm

This defines a Markov chain of transition matrix:

$$P_{i,j} = \begin{cases} 1/\deg(j) & \text{if there is a link } j \to i \\ 0 & \text{otherwise,} \end{cases}$$

- After a long time, the surfer is more likely to be on an *important webpage*.
- If μ is the invariant measure of P (provided P verifies the hypotheses of Perron-Frobenius), we take

$$\mu_i \;=\; \text{« importance of webpage } i \text{ ».}$$

PageRank Algorithm

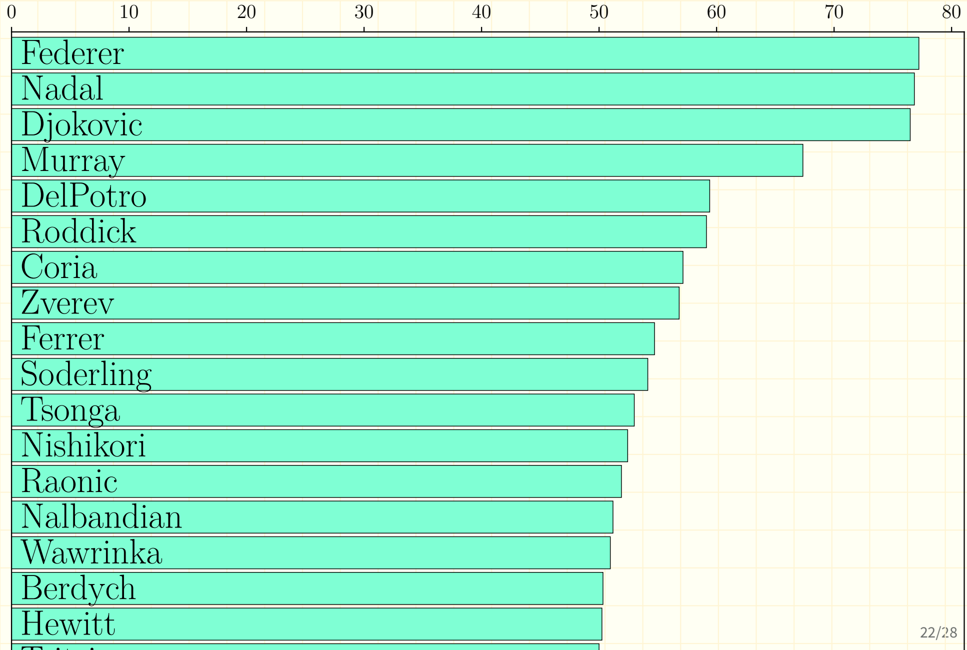
Application: ranking Tennis players

Goal: rank the following players:

Federer, Nadal, Djokovic, Murray, Del Potro, Roddick, Coria, Zverev, Ferrer, Soderling, Tsonga, Nishikori, Raonic, Nalbandian, Wawrinka, Berdych, Hewitt, Tsitsipas, Monfils, Gonzalez, Thiem, Ljubicic, Davydenko, Cilic, Pouille, Safin, Isner, Dimitrov, Medvedev, Ferrero, Goffin, Bautista Agut, Sock, Gasquet, Simon, Blake, Monaco, Coric, Stepanek, Khachanov, Almagro, Robredo, Verdasco, Anderson, Youzhny, Baghdatis, Dolgopolov, Kohlschreiber, Fognini, Melzer, Paire, Querrey, Tomic, Basilashvili.

Data: Head-to Head records (number of times that player x has defeated player y)

Ranking by % of victories



The random spectator

Imagine the following « random spectator »:

- At time t, the spectator believes that player j is the best: $X_t = j$.
- Then, he picks a game of player j uniformly at random:
  - if player j wins, then the spectator still believes that j is the best: $X_{t+1} = j$.
  - otherwise, the spectator changes his mind and now believes that player i who defeated j is the best: $X_{t+1} = i$.

The random spectator

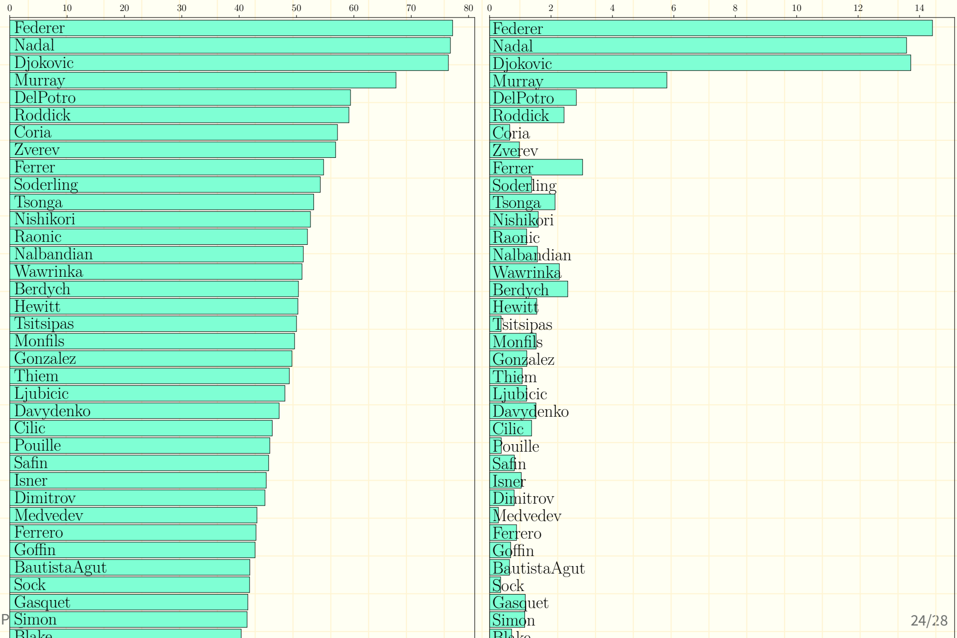
Imagine the following « random spectator »:

- At time t, the spectator believes that player j is the best: $X_t = j$.
- Then, he picks a game of player j uniformly at random:
  - if player j wins, then the spectator still believes that j is the best: $X_{t+1} = j$.
  - otherwise, the spectator changes his mind and now believes that player i who defeated j is the best: $X_{t+1} = i$.

This defines a transition matrix P. We rank the players according to the stationary distribution μ of

$$M = \alpha P + \frac{1-\alpha}{N} J$$

Naive ranking vs PageRank



The Spectral Theorem

The spectral theorem

Theorem

Let $A \in \mathbb{R}^{n\times n}$ be a **symmetric** matrix. Then there is a orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of A.

The spectral theorem

Matrix formulation

Theorem

Let $A \in \mathbb{R}^{n\times n}$ be a **symmetric** matrix. Then there exists an orthogonal matrix P and a diagonal matrix D of sizes $n \times n$ such that

$$A = PDP^{\mathsf{T}}.$$

Questions?

Questions?