MATH466/MATH766
Math of machine learning
03/26 Lecture 18 Random Walk on Graphs and PageRank

References:

- Convergence theorem for finite Markov chains by Ari Freedman
  https://math.uchicago.edu/~may/REU2017/REUPapers/Freedman.pdf
- Wikipedia PageRank https://en.wikipedia.org/wiki/PageRank
- DKU Math405 Spring21 week2 lecture notes by Prof. Xiuyuan Cheng

Todays contents:

- Random walk on graphs and random walk matrix
- Stationary distribution of random walks and PageRank

Important concepts:

- random walk matrix
- stationary distribution
- ergodic

Recommend reading:

-

Warm up.

1. $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A) = r$, $\dim(\mathrm{null}(A)) =$ ________

   $\dim(\mathrm{null}(A^T)) =$ ________

$\Rightarrow$ $A \in \mathbb{R}^{n\times n}$, λ is an eigenvalue of A,

$$\dim(\mathrm{null}(A-\lambda I)) = \dim(\mathrm{null}(A^T-\lambda I))$$

2. $\Pr[A,B] = \Pr[B|A]\Pr[A]$

   If $A|C$, $B|C$ independent then $\Pr[A,B|C] = \Pr[A|C]\Pr[B|C]$

Random Walk and Random Walk Matrix

- Def. (random process) A random process is a collection of random variables $\{X_t : t \in T\}$, T can be discrete or continuous time

  e.g. (X_1, X_2, X_3), $(X_1, X_2, X_3, X_4, \cdots)$,

  $\{X_t : t \in [0,1]\}$, $\{X_t : t \in [0,+\infty)\}$

- Def (Markovian) A discrete time random process $\{X_t, t=0,1,2,\cdots\}$ is Markovian if $\Pr[X_{t+1} = x \mid X_0 = x_0, \cdots, X_t = x_t] = \Pr[X_{t+1} = x \mid X_t = x_t]$

  (means: conditioning on current state, the future and past are independent)

  (Markovian random process is also called Markov chain)

To describe a Markov chain, need $\Pr[X_{t+1} = x \mid X_t = y]$ for any x, y, t

- Def (stationary / homogenous) A Markov chain is stationary if for any t, $\Pr[X_{t+1} = y \mid X_t = x] = \Pr[X_1 = y \mid X_0 = x]$

  (transition probability independent of t)

- Def (Markov matrix) $P \in \mathbb{R}^{n \times n}$ is a Markov matrix if

  $$P_{ij} \geq 0 \quad \text{and} \quad \sum_{j=1}^{n} P_{ij} = 1 \quad \text{for each } i$$

  (P_{ij} = at node i, the probability of jumping to node j (in 1 step))

- Prop. If P is a Markov matrix,

  then for any $k \in \mathbb{N}$, P^k is also a Markov matrix

  ($(P^k)_{ij}$ = at node i, the probability of jumping to node j in k steps)

- Prop. $G = (V, E)$ w/. affinity matrix W, degree matrix D

  $P := D^{-1} W$ is a Markov matrix

  (HW: P is diagonalizable; $\lambda(P) \in [-1, 1]$)

Stationary Distribution

Consider $\pi \in \mathbb{R}^n$, $\pi_i \geq 0$, $\sum_i \pi_i = 1$.

π is a probability distribution on the graph G

- Prop. P Markov matrix, π probability distribution

  then $P^T\pi$ is also a probability distribution

  ($P^T\pi$: initial state π, the state after 1 jump)

- Def : (stationary distribution) $\pi = P^T\pi$

  (eigenvector of P^T corresponding to eigenvalue 1)

- Prop : $\pi = \frac{1}{\sum_i d_i}\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$ is a stationary distribution of $P = D^{-1}W$

(If π is the unique stationary distribution of a Markov chain P
then π_i reflects the "importance" of the node i.)

Properties of Stationary Distribution

① uniqueness

- Def (Irreducible) A Markov chain is irreducible if

  for any i, j, there exists a $t > 0$ s.t. $(P^t)_{ij} > 0$

  (In our graph setting, irreducible $\Leftrightarrow$ the graph is connected)

- Prop. If P is irreducible and has a stationary distribution π

  then π is the unique stationary distribution

Pf for graph induced P : Irreducible $\Rightarrow$ G is connected

$\Rightarrow \dim(\mathrm{null}(P - I)) = 1 \Rightarrow \dim(\mathrm{null}(P^T - I)) = 1 \Rightarrow$ uniqueness

Pf for general Markov matrix P : (check ref for details)

Irreducible $\Rightarrow$ $\dim(\text{null}(P-I)) = 1$ $\Rightarrow$ ....

② Existence :

Q: Why not use $\dim(\text{null}(P^T - I)) = \dim(\text{null}(P-I)) \geq 1$

to argue the existence ?

Consider a Markov chain. define

the hitting time $\tau_x := \min\{t \geq 0 : X_t = x\}$

the first return time $\tau_x^+ := \min\{t > 0 : X_t = x\}$

$$\mathbb{E}_x[*] := \mathbb{E}[* \mid X_0 = x]$$

Lemma. If P is irreducible, then $\mathbb{E}_x[\tau_y^+]$ is finite for any x, y

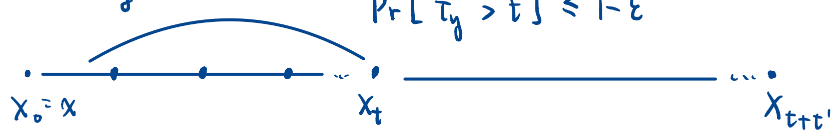
Proof outline : ① $\mathbb{E}_x[\tau_y^+] = \sum_{k=1}^{+\infty} k \Pr[\tau_y^+ = k]$

$$= \sum_{k=1}^{+\infty} \Pr[\tau_y^+ = k] + \sum_{k=2}^{+\infty} \Pr[\tau_y^+ = k] + \cdots$$

$$= \sum_{k=0}^{+\infty} \Pr[\tau_y^+ > k]$$

② If $(P^t)_{xy} = \varepsilon > 0$

$\Pr[\tau_y^+ > t] \leq 1-\varepsilon$



$\Pr[\tau_y^+ > t+t'] \leq (1-\varepsilon)(1-\varepsilon')$

depends on X_t

$$\Pr[\tau_y^+ > a] \leq \Pr[\tau_y^+ > a-c] \quad (c > 0)$$

③ b/c finite state, $\exists\, r, \varepsilon > 0$

s.t. $\Pr[\tau_y^+ > r \mid X_0 = x] \leq 1-\varepsilon$ for any x, y

$$\Rightarrow \mathbb{E}_x[\tau_y^+] \leq r \sum_{k=0}^{+\infty} (1-\varepsilon)^k$$

Thm. If P is irreducible, then it has a unique stationary distribution π with $\pi(x) > 0 \ \forall x$, given by $\pi(x) = \dfrac{1}{\mathbb{E}_x[\tau_x^+]}$

Intuition: When $P = D^{-1}W$, $\pi(x) = d(x) / \sum_y d(y)$,

$\pi(x) \sim$ vol of x $\sim$ the amount of connections to x

$\sim$ the avg # of visit to x in a Markov chain

Proof outline:

① Fix any state z

define $\tilde{\pi}_z(y) := \mathbb{E}_z[\#$ of visit to y before returning to $z]$

$$= \sum_{k=1}^{+\infty} k \Pr_z[X_{t_1} = X_{t_2} = \dots = X_{t_k} = X_{t_{k+1}} = y, t_k + 1 \le \tau_z^+ < t_{k+1}]$$

$$= \sum_{t=0}^{+\infty} \Pr_z[X_t = y, \tau_z^+ > t]$$

- irreducible $\Rightarrow \tilde{\pi}_z(y) > 0$
- $\tilde{\pi}_z(y) \le \mathbb{E}_z[\tau_z^+]$, b/c Lemma, $\tilde{\pi}_z(y) < \infty$
- prove that $\tilde{\pi}_z$ is stationary

$$(P^T \tilde{\pi}_z)(y) = \sum_x P(x,y)\tilde{\pi}(x) \quad < \infty, \quad P(x,y)\tilde{\pi}(x) > 0$$

$$= \sum_x \sum_{t=0}^{+\infty} \Pr_z[X_t = x, \tau_z^+ \ge t+1] P(x,y)$$

(nontrivial) $$= \sum_{t=0}^{+\infty} \sum_x \underline{\Pr_z[X_t = x, \tau_z^+ \ge t+1] P(x,y)}$$

$\Pr_z[X_t = x, \tau_z^+ \ge t+1] P(x,y)$

$= \Pr_z[X_t = x, \tau_z^+ \ge t+1] \Pr_z[X_{t+1} = y \mid X_t = x]$

$= \Pr_z[\underline{\tau_z^+ \ge t+1 \mid X_t = x}] \Pr_z[\underline{X_{t+1} = y \mid X_t = x}] \Pr_z[X_t = x]$

independent *

* b/c the prob. of the event $\tau_z^+ \ge t+1$ only depends on history $X_1, X_2, \dots, X_t$ and is independent of future $X_{t+1}, X_{t+2}, \dots$

$= \Pr_z[X_{t+1} = y, \tau_z^+ \ge t+1 \mid X_t = x] \Pr_z[X_t = x]$

$= \Pr_z[X_t = x, X_{t+1} = y, \tau_z^+ \ge t+1]$

$$\sum_x \Pr_z[X_t = x, X_{t+1} = y, \tau_z^+ \geq t+1] = \Pr_z[X_{t+1} = y, \tau_z^+ \geq t+1]$$

$$\Rightarrow (P^T \tilde{\pi}_z)(y) = \sum_{t=1}^{+\infty} \Pr_z[X_t = y, \tau_z^+ \geq t]$$

$$= \sum_{t=1}^{+\infty} \Pr_z[X_t = y, \tau_z^+ > t] + \sum_{t=1}^{+\infty} \Pr_z[X_t = y, \tau_z^+ = t]$$

$$= \tilde{\pi}_z(y) \underbrace{- \Pr_z[X_0 = y, \tau_z^+ > 0] + \sum_{t=1}^{+\infty} \Pr_z[X_t = y, \tau_z^+ = t]}$$

case (i) $y \neq z$, $= -0 + \sum_{t=1}^{+\infty} 0 = 0$

case (ii) $y = z$, $= -\Pr_z[\tau_z^+ > 0] + \sum_{t=1}^{+\infty} \Pr_z[\tau_z^+ = t] = 0$

$$= \tilde{\pi}_z(y)$$

② "Normalize $\tilde{\pi}_z$"

- $\sum_x \tilde{\pi}_z(x) = \sum_x \mathbb{E}_z[\#\text{ of visits to } x \text{ before returning to } z]$

  (nontrivial) $= \mathbb{E}_z \sum_x [\#\text{ of visits to } x \text{ before returning to } z]$

  $= \mathbb{E}_z[\tau_z^+]$

- $\pi(x) := \dfrac{\tilde{\pi}_z(x)}{\mathbb{E}_z(\tau_z^+)}$ well-defined b/c $\mathbb{E}_z(\tau_z^+) > 0$

  $\pi(x)$ is a stationary distribution.

  By Lemma, $\pi(x)$ is unique, thus invariant to z

- Take $z = x$, $\pi(x) = \dfrac{1}{\mathbb{E}_x(\tau_x^+)}$

PageRank

step (1) Let $P = (1-\delta) D^{-1} W + \delta \frac{1}{n} 1_n 1_n^T$, $\delta \in (0,1)$

irreducible!

step (2) compute stationary distribution

approach ① Monte-Carlo simulation:

simulate random walks on the graph

and compute long-time average of times of visits

approach ② power method $(P^T)^t \pi_0$, any π_0, large t

(requires the MC to be aperiodic as well)

not oscillating in a fixed cycle