Foundations of Data Science & Machine Learning

Summary — Week 08
Devansh Singh Rathore
111701011

B.Tech. in Computer Science & Engineering
Indian Institute of Technology Palakkad

April 19, 2021

Abstract

This week's lectures discusses random walks on directed graphs and its properties. Further we discuss stationary distributions of random walks with its real life applications.

1 Random Walks on Graphs

Definition: A **Random Walks** consists of a 'walk' on graph where the destination is chosen according to the current state and outgoing path probabilities. The sum of probabilities of outgoing paths from a state is equal to 1.

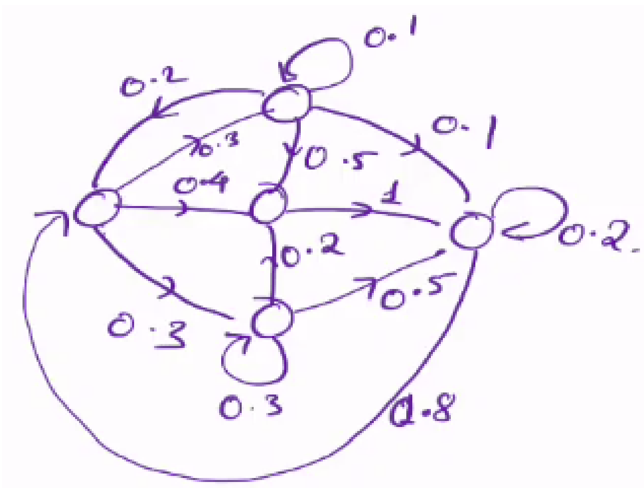


Fig 1.0 A sample Graph

→ A good example for random walk can be (web pages, hyperlinks).

Fig 1.1 Weight of edge representing the probability P_{ij}

→ Edge weights = Transition probability

$$P_{ij} = P[Next\,State = j/Current\,State = i]$$

$$Hence\,\forall i \in V(G), \sum_{j \in V(G)} P_{ij} = 1$$

→ Transition Probability Matrix (TPM):

$$P = \begin{bmatrix} P_{11} & ... & P_{1n} \\ P_{21} & ... & P_{2n} \\ : & & \\ P_{n1} & ... & P_{nn} \end{bmatrix}_{n \times n}$$

$$\begin{bmatrix} P_{11} & ... & P_{1n} \\ P_{21} & ... & P_{2n} \\ : & & \\ P_{n1} & ... & P_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ : \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ : \\ 1 \end{bmatrix}$$

→ P1 ≡ Every row of P sums to 1.
→ P is not symmetric as it's not necessary that $P_{ij} = P_{ji}$
→ Hence 1I = $[1, 1, ..., 1]^T$ is a (right) eigen vector of P with eigen value 1.
In fact, if P is real non-negative matrix then,
P is a stochastic matrix <=> P 1I = 1I

Corollary: If P_1 and P_2 are n × n stochastic matrices, then so is P_1P_2

$$P_1P_21| = P_1(P_21|) = P_11| = 1|$$

Corollary: If P is a stochastic matrix, then $\forall k \in \mathbb{N}, P^k$ is a stochastic matrix. (proof using previous corollary)

$$P^k[i, j] = P[\text{State after k steps = j / current step = i}]$$

Proposition: All eigen values of P have magnitude ≤ 1. i.e. Px = $\lambda x \Rightarrow |\lambda| \leq 1$
Proof: Let x = $(x_1, x_2, ...x_n)$ be a eigen vector corresponding to λ s.t. some $|x_j| \leq 1 \forall j$

$$Px = \lambda x$$

$$< (i^{th}\,row\,of\,P), x >= \lambda x_i$$

$$\sum_{j=1}^{n} P_{ij}x_j = \lambda 1$$

$$|\lambda| \leq \sum_{j=1}^{n} P_{ij}|x_j| \leq \sum_{j=1}^{n} P_{ij}1 = 1$$

→ What is multiplicity of $\lambda = 1$?
If $\lambda = 1$, then $x_j = 1 \forall j$ s.t $P_{ij} \neq 0$
If every vertex is reachable from i, then $x_j = 1 \forall j \Rightarrow x = 1I$.

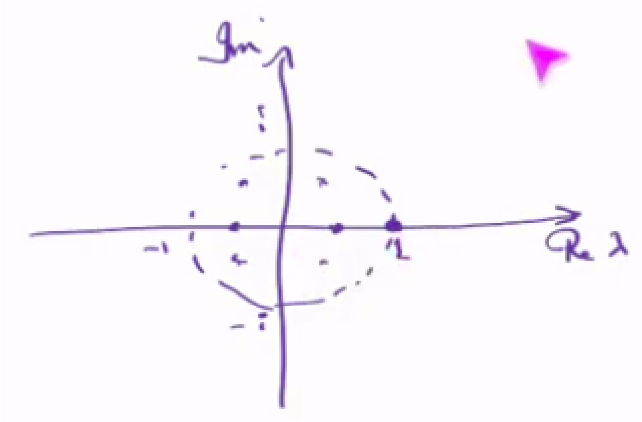


Fig 1.2 Eigen value

Proposition: If the embedding graph G is strongly connected then the eigen value 1 has the multiplicity 1.

$\equiv$ 1I is the unique eigen vector (upto scaling) for $\lambda = 1$.

$\rightarrow$ Converse? (Exercise)

(Hint: Sink component)

$\rightarrow$ When do we get $\lambda = -1$?

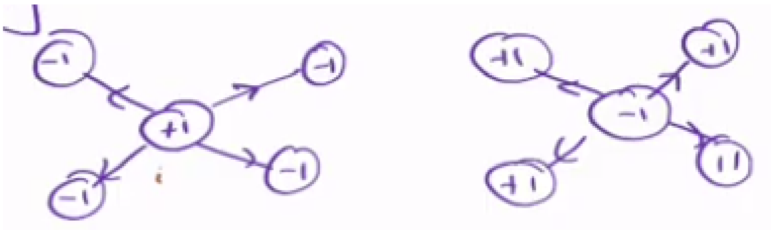


Fig 1.3 Graph in $\lambda = -1$ case

This could happen when G is bipartite graph

Strongly connected $\Rightarrow$ -1 also has multiplicity 1.

$\rightarrow$ Is P full rank?

Not necessarily.

A possible case when P is full rank:

$$\begin{bmatrix} 1/3 & 1/3... & 1/3 \\ 1/3 & 1/3... & 1/3 \\ \vdots & & \\ 1/3 & 1/3... & 1/3 \end{bmatrix}$$

Interestingly the n $\times$ (n+1) matrix P - I : 1I has rank n. (exercise)

$\rightarrow$ In an undirected graph case, $P_{ij} = P_{ji}$ and the P is symmetric matrix.

$\rightarrow$ P^T? ... It capture "evolution of random walk"

P^T is more important than P. Let p = $(p_1, p_2, ..., p_n)$ be the node probabilities at current step. Then, $p' = P^T p$ gives the node probability at the next step.

$$p'_i = \sum_{j=1}^{n} P_{ji} p_j = [P^T p]_i$$

$$p(t+1) = P^T p(t)$$

$\rightarrow$ $x = (x_1, x_2, ..., x_n)$ is called probability vector if $x_i \geq 0 \, \forall i$ and $\sum_{i=1}^{n} x_i = 1$
$\rightarrow$ Properties of P^T:

- Eigen values (and multiplicities) pf P and P^T are the same. (Proof using determinents: $\det(P^T - \lambda I) = \det((P - \lambda I)^T) = \det((P - \lambda I)))$.

- x is a prob. vector $\Rightarrow$ P^T x is a prob. vector

  Proof: Let $y = P^T x$, $y_i \geq 0$ obviously

$$\sum_{i=1}^{n} y_i = 1|^T y = 1|^T P^T x = (P\,1|)^T x = (1|)^T x = 1|$$

- P^T also has a unique eigen vector (upto scaling) for $\lambda = 1$. (considering strongly connected graph for 'unique'). But is it a prob. vector? YES!

2 Stationary Distributions of Random Walks

Definition: $\pi \in \mathbb{R}^n$ is called a **Stationary Distribution** of a random walk with transition matrix P if:
(i) π is a prob. vector &
(ii) $P^T \pi = \pi$

Theorm 1: (Fundamental theorm of finite Markov Chain)
$\rightarrow$ A random walk on every finite graph has a stationary distribution.
Observation: Finiteness is necessary.
Standard Proof: Key idea: **"Fixed Point Theorems"**
Let $f : X \rightarrow X$, then a point $x \in X$ s.t. f(x) = x is called a **fixed point** of f.
Example: Any continuous function from a closed interval [a,b] to itself has a fixed point.
Let f : [a,b] $\rightarrow$ [a,b], continuous

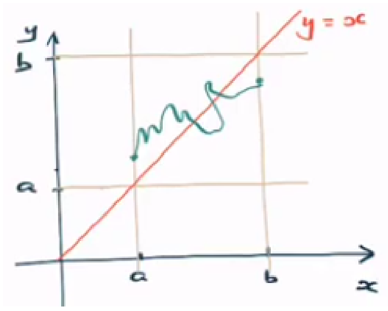


Fig 2.0 function f

Since f is continuous function, f will intersect with function y = x at at least one point. Hence there will be at least one fixed point independent of what f is.
Fact 1: If X is closed a compact convex set and f is continuous, then f has a fixed point. (Brouwer's FP theorm)
Fact 2: Closed and bounded sets in $\mathbb{R}^n$ are compact.
$\rightarrow$ What's our X?
X = set of all prob. vectors in $\mathbb{R}^n$. $(n = |V(G)|)$
$= \{(p_1, .., p_n) : p_i \geq 0, \sum p_i = 1\}$

→ Verify that X is:

(i) convex ($p, q \in X$, $then\, \lambda p + (1-\lambda)q \in X$)

(ii) bounded ($||x||_\infty : max x_i$, $then \{||x||_\infty : x \in X\} \leq B = 1$)

(iii) closed

→ Then f : X → X

x $\mapsto P^T x$

→ Verify f is continuous.

So f has a fixed point in X. That is your stationary distribution.

Direct Proof:

Let x be any prob. vector

Consider the sequence of prob. vector - $x, xP, xP^2, xP^3, ...$ (Evolution of the random walk x(0), x(1), ... where $x(t) = xP^t$)

Let **"long-term average"** be a(t) = (1/t)(x(0) + x(1) + ... x(t-1))
$= (1/t)(x + xP + ... + xP^{t-1})$

Claim 1: $\forall t$, a(t) is a prob. vector

therefore, a(1), a(2), ... $\in X$

X is bounded ⇒ a(1), a(2), ... contains a convergent subsequence $a(t_1), a(t_2), ...$(proof: pigeonhole principle)

X is closed ⇒ $lim_{n \to \infty} a(t_n) = a \in X$. (i.e. a is a prob. vector)

$$\forall t,\, a(t)P - a(t) = (1/t)(xP + xP^2 + ... + xP^t) - (1/t)(x + xP + ... + xP^{t-1})$$

$$= (1/t)(xP^t - x)$$

$$||a(t)P - a(t)||_\infty \leq (1/t) \quad - (*)$$

$$||a(t_n)P - a(t_n)||_\infty \leq (1/t_n) \to 0\, as\, n \to \infty$$

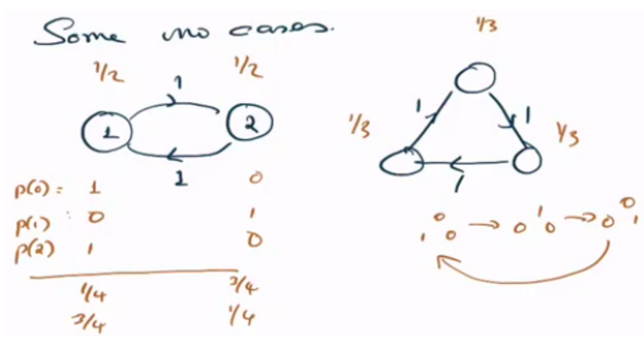
$$\Rightarrow ||aP - a||_\infty = 0$$

$$aP = a$$

Hence a is a stationary distribution.

Question. If p(0) is an arbitrary starting distribution. Does $p(n) = p(0)P^n \to \pi\ as\ n \to \infty$ **?**

Ans. Not always but Yes in most cases.



(Fig 2.1)

Yes iff gcd of lengths of all directed cycles in G is 1.

2.1 Summary

1. Every random walk on a finite graph has a stationary distribution π. $(\pi P = \pi)$
2. If the graph is strongly connected then the stationary distribution is unique and $a(t) \to \pi \; as \; t \to \infty$
3. If the graph G is strongly connected and gcd(cycle length) = 1, for any prob. vector p(0), $p(t) \to \pi \; as \; t \to \infty$

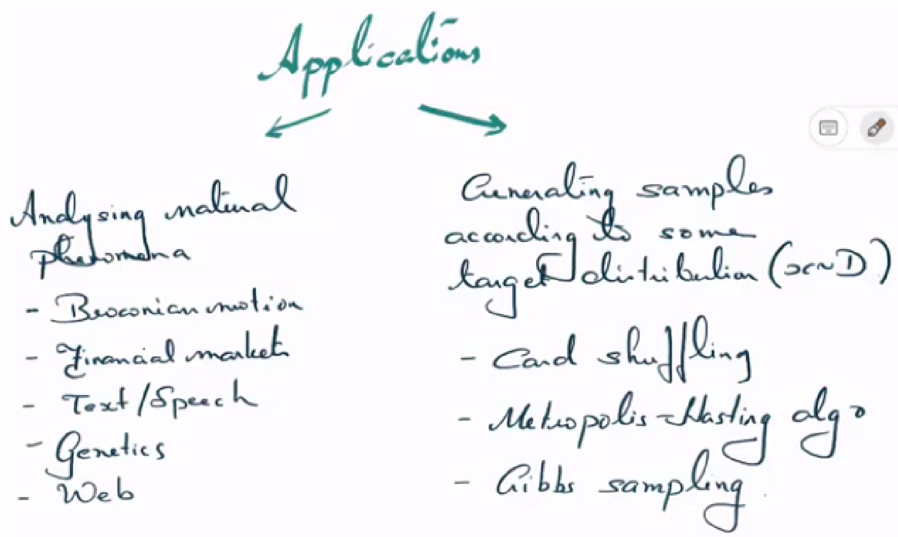


Fig 2.2 applications of point(3.)