Mathematics for Machine Learning (AI 512)

Amit Chattopadhyay

IIIT-Bangalore

M3: Random Walk

Syllabus

Topics

M3: Markov Chains, Stationary Property and Random Walk
M3:. Page Rank Algorithm
M3. Markov Chain Monte Carlo (MCMC) Sampling
M4. Multivariate Distributions, K-Means
M4. EM Algorithm
M4. EM Algorithm for GMM.

Reference Books

1. Foundations of Data Science. By John Hopcroft, Ravindran Kannan.
2. Finite Markov Chains and Algorithmic Applications. By OLLE HÄGGSTRÖM.
3. Introduction to Probability Models. By S.M. Ross.
4.Pattern Recognition and Machine Learning. By Christopher M. Bishop.
5. Mathematics for Machine Learning. By Marc Peter Deisenroth et al.

Summary: What we have seen in Markov Chains

- $(X_0, X_1, \ldots)$: a **Markov chain** (MC) with **state space** $S = \{s_1, s_2, \ldots, s_N\}$ and **transition Matrix** $\mathbb{P}$
- **Initial distribution** of the chain, i.e. distribution of X_0:

$$\underline{p}^{(0)} = [p_1^{(0)}, p_2^{(0)}, \ldots, p_N^{(0)}]$$

with $p_i^{(0)} \geq 0$ and $\sum_{i=1}^N p_i^{(0)} = 1$

- **Transition probabilities**: $p_{i,j} = P(X_{n+1} = s_j | X_n = s_i)$ and **transition matrix**

$$\mathbb{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,N} \end{bmatrix}$$

with $\sum_{i=1}^N p_{i,j} = 1$.

Summary: What we have seen in Markov Chains

- **Distribution of X_n:** $\underline{p}^{(n)} = [p_1^{(n)}, p_2^{(n)}, \ldots, p_N^{(n)}]$ where

$$\boxed{\underline{p}^{(n)} = \underline{p}^{(0)}\mathbb{P}^n}$$

- ***n*-step transition probabilities**:

$$P(X_{m+n} = s_j | X_m = s_i) = (\mathbb{P}^n)_{i,j}$$

  for $m \geq 0, n \geq 0$

- **Transition or State graph:** pictorial representation of MC

- **Goal:** To understand the conditions on $\mathbb{P}$ so that the limiting distribution $\lim_{n\to\infty} \underline{p}^{(n)}$ exists uniquely which is the **stationary distribution** $\pi = [\pi_1, \pi_2, \ldots, \pi_N]$, i.e.

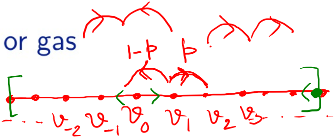
$$\boxed{\pi = \pi\mathbb{P}}$$

.

Random Walk

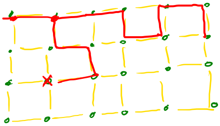
Random Walk

- A type of Stochastic Process - in particular, a Markov Chain e.g.
  - path traced by a molecule in a liquid or gas
  - search path of a foraging animal
  - price of a fluctuating stock etc.



- Describes a **path**: succession of random steps on some "**mathematical space**"

- **Mathematical space:**
  - Real line ($\mathbb{R}^1$)
  - Integer number line $\mathbb{Z}$
  - *d*-dimensional integer lattice $\mathbb{Z}^d$
  - Graphs : Undirected or directed

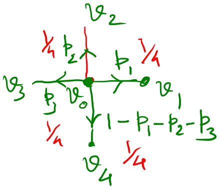


v_1, v_2 . . .

n

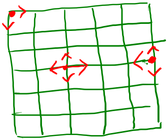
Random Walk on Regular lattice

At each step –jump to ‘another’ site according to some probability distribution



Random Walk on Regular Lattice

- **Simple Random Walk:** The random walker can only jump to "**neighbouring**" sites of the lattice
- **Simple Symmetric Random Walk:** The probabilities of the random walker jumping to immediate"**neighbours**" are the same.
- **Simple Bordered Symmetric Random Walk:**
  - state space is limited to a finite domain
  - transition probabilities depend on the location of the state;
  - on **margin** and **corner states** the movement is limited.



Random Walk on a Graph

Graph: $G = (V, E)$, V : set of vertices, E : set of edges

- **Walk:** an alternating sequence $\langle v_1, e_1, v_2, e_2, \ldots, e_{N-1}, v_N \rangle$ of vertices and edges;
  Start vertex: v_1, **End vertex:** v_N and **edge:** $e_i = (v_i, v_{i+1}) \in E$, $\forall$ $i = 1, 2, \ldots, N-1$.

- **Random Walk:** a sequence of vertices generated from a start vertex by selecting an edge **randomly** (following some rule) and repeating the process.

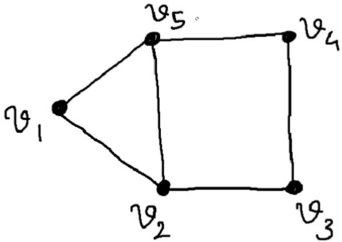
- **Goal:** Under what conditions, the fraction of time the random walk spends at various vertices converges to a **stationary probability**?

$$\underline{p} = (p_1, p_2, \ldots, p_k)$$

$$p_1 = \frac{n_1}{N}, \quad p_2 = \frac{n_2}{N}, \quad \ldots, \quad p_k = \frac{n_k}{N}$$

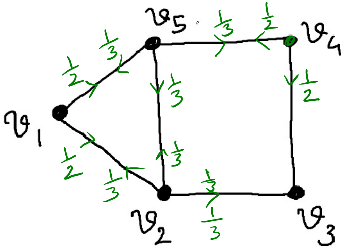
Examples

Example: Random walks on a graph $G = (V, E)$



$V = \{v_1, v_2, v_3, v_4, v_5\}$ and
$E = \{\langle v_1, v_2\rangle \langle v_2, v_3\rangle \langle v_3, v_4\rangle \langle v_4, v_5\rangle \langle v_5, v_2\rangle, \langle v_1, v_5\rangle\}$.
Two vertices are said to be **neighbours** if they share an edge.

A **random walk** on $G = (V, E)$ is a **Markov chain** with the state space
$V = \{v_1, v_2, v_3, v_4, v_5\}$ and



$$\mathbb{P} = \begin{array}{c|c} & \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \end{bmatrix} \end{array}$$

transition probabilities:

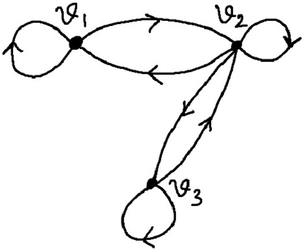
$$p_{i,j} = \begin{cases} \frac{1}{d_i} & \text{if } v_i \text{ and } v_j \text{ are neighbours} \\ 0 & \text{otherwise} \end{cases}$$

where d_i is the number of neighbours (**degree**) of a vertex v_i.

(*probability of the RW at vertex v_i at time n, selecting an edge $\langle v_i, v_j \rangle$ to* move to vertex v_j at $n+1$)

Random Walk on a Directed Graph

- **Directed graph:** $G = (V, E)$; V : set of vertices, E : **ordered pair** of vertices



$$\mathbb{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & 0 \\ p_{2,1} & p_{2,2} & p_{2,3} \\ 0 & p_{3,2} & p_{3,3} \end{bmatrix}$$

$V = \{v_1, v_2, v_3\}$ and
$E = \{(v_1, v_1), (v_1, v_2), (v_2, v_1), (v_2, v_2), (v_2, v_3), (v_3, v_2), (v_3, v_3)\}$.

- **State space:** $V = \{v_1, v_2, v_3\}$; **Transition matrix:** $\mathbb{P}$ (rule to be found depending on application)

Random Walk: Initial Distribution

- **Initial distribution:** If RW starts at vertex v_1, $\underline{\mathrm{p}}^{(0)} = [1, 0, 0]$.

  In general, $\underline{\mathrm{p}}^{(0)} = [p_1^{(0)}, p_2^{(0)}, p_3^{(0)}]$, $p_i^{(0)} \geq 0$ for $i = 1, 2, 3$ and $p_1^{(0)} + p_2^{(0)} + p_3^{(0)} = 1$.

Q: What does it mean that a RW start at v_i with probability $p_i^{(0)}$?

Ans. Consider multiple (say, M) **realizations** of the RW and its **ensemble**. Let n_i be the number of realizations with start vertex v_i.

$$p_i^{(0)} = \lim_{M \to \infty} \tilde{p}_i = \lim_{M \to \infty} \frac{n_i}{M}$$

Random Walk: Initial Distribution

Q: Given an initial distribution: $\underline{\mathbf{p}} = [p_1, p_2, \ldots, p_N]$, what does it mean that a RW start with a vertex 'follows' $\underline{\mathbf{p}}$?

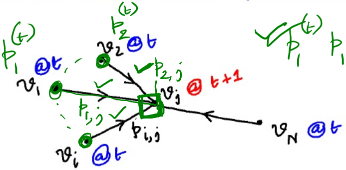
Ans. In stable situation, RW gives a **Sampling** from the distribution of $\underline{\mathrm{p}}$

(will be discussed in **MCMC sampling**)

$$\boxed{\underline{p}^{(t+1)} = \underline{p}^{(t)}\,\mathbb{P}} \qquad \boxed{p^{(n)} = p^{(0)}\,\mathbb{P}^{n}}$$

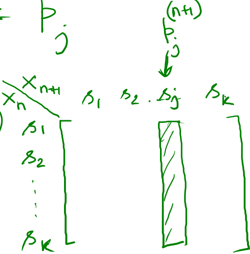
- $\mathbb{P} = (p_{i,j})_{N\times N}$;

  $p_{i,j}$: probability of the RW at vertex v_i to select the edge to vertex v_j



$$p_1^{(t)}\,p_{1,j} + p_2^{(t)}\,p_{2,j} + \cdots + p_N^{(t)}\,p_{N,j} = p_j^{(t+1)}$$

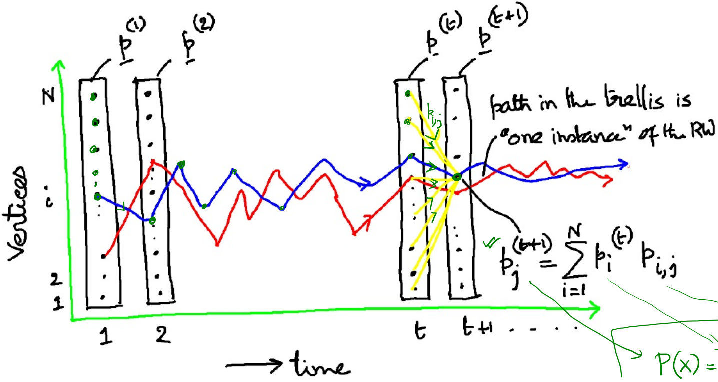
$$\underline{(p_1^{(n)}, p_2^{(n)}, \dots, p_k^{(n)})}$$



- probability of being a vertex v_j at $t+1$:

  $\boxed{p_j^{(t+1)} = \sum_{i=1}^{N} p_i^{(t)}\, p_{i,j}}$ $(j = 1, 2, \dots, N)$
- $\boxed{\underline{p}^{(t+1)} = \underline{p}^{(t)}\mathbb{P}}$ (for $t = 0, 1, 2, \dots$)

Random Walk: Trellis View



$P(X) = P(A_1)P(X|A_1) + \cdots$
$+ P(A_N)P(X|A_N)$

- Path in the trellis is a realization of the random walk
- interpretation of $p_j^{(t+1)} = \sum_{i=1}^{N} p_i^{(t)} p_{i,j}$
- Limiting probability: Stationary distribution (**if exists**)