Mathematics for Machine Learning (AI 512)

Amit Chattopadhyay

IIIT-Bangalore

Lecture: Stationary Distribution

Stationary Distribution

A row vector $\pi = (\pi_1, \ldots, \pi_N)$ is said to be a **stationary distribution** for the MC, if it satisfies:

(i) $\pi_i \geq 0$ for $i = 1, 2, \ldots, N$ and $\sum_{i=1}^{N} \pi_i = 1$ (p.m.f.)

(ii) $\pi \mathbb{P} = \pi$

Examples:

1. Every probability distribution on the states is a stationary probability distribution when $\mathbb{P}$ is the identity matrix.
2. If $\mathbb{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{10} & \frac{9}{10} \end{bmatrix}$ then $\pi = (\frac{1}{6}, \frac{5}{6})$ satisfies $\pi\mathbb{P} = \pi$.

$\underline{p}^{(0)} \mathbb{P} = \underline{p}^{(1)}$

$\underline{p}^{(1)} \mathbb{P} = \underline{p}^{(2)}$

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

Lemma: The largest eigenvalue of a **row-stochastic** square matrix $\mathbb{P}$ is 1.

Proof: Let $\mathbb{P}$ be an $N \times N$ row-stochastic matrix.

Then $\boxed{\mathbb{P}\underline{1} = \underline{1}} \implies \underline{1}$ is a right eigenvector with eigenvalue 1.

$$\mathbb{P}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

To show for any other eigenvalue λ of $\mathbb{P}$, $|\lambda| \leq 1$:

$$\mathbb{P} = (p_{ij})_{N\times N} \qquad \boxed{\mathbb{P}\underline{v} = \lambda \underline{v}} \qquad \text{with } \underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}$$

$\exists$ index k s.t. $|v_k| \geq |v_i| \ \forall i = 1, 2, \ldots, N$

Comparing the k-th entry

$$\sum_{j=1}^{N} p_{kj} v_j = \lambda v_k \Rightarrow |\lambda||v_k| = \left|\sum_{j=1}^{N} p_{kj} v_j\right| \leq \sum_{j=1}^{N} p_{kj} |v_j| \leq \sum_{j=1}^{N} p_{kj} |v_k| = |v_k|$$

$$\Rightarrow |\lambda| \leq 1.$$

(P1) If $(\lambda, \mathbf{v})$ be a right eigen-pair of square matrix $\mathbb{P}$, then $(\lambda, \mathbf{v}^T)$ is a left eigen-pair of the matrix $\mathbb{P}^T$.

$P\underline{v} = \lambda \underline{v}$

$\Rightarrow \underline{v}^T \mathbb{P}^T = \lambda \underline{v}^T$

(P2) A matrix $\mathbb{P}$ and its transpose $\mathbb{P}^T$ have the same set of eigenvalues.

Perron-Frobenius (P-F) Theorem

1 (if aperiodic MC)

Let $\mathbb{A}$ be an irreducible, non-negative $N \times N$ matrix with period p and spectral radius $\rho(\mathbb{A}) = r$. Then

$= \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_N|\}$

1. The number r is a **unique eigenvalue** of $\mathbb{A}$ (it is a simple root of the characteristic equation of $\mathbb{A}$).

   $\psi(\lambda) = \det(A - \lambda I) = 0$

2. $\mathbb{A}$ has a left eigenvector $\mathbf{v}$ with associated eigenvalue r, and $\mathbf{v}$ has all positive entries.

   $|\lambda_1| = |\lambda_2| = \quad = |\lambda_p| = r$

3. $\mathbb{A}$ has exactly p complex eigenvalues with modulus r and each is a simple root of the characteristic polynomial of $\mathbb{A}$.

Note: Using P-F theorem, an irreducible and aperiodic transition matrix $\mathbb{P}$ (which is row-stochastic) has a unique eigenvalue 1 and all other eigenvalues have moduli **strictly less** than 1.

Uniqueness and Convergence

Theorem: (i) An **irreducible**, **aperiodic**, **homogeneous finite** Markov chain has a **unique stationary distribution**. (Ref. Olle Häggström)

(ii) Furthermore, if $\mathbb{P}$ is **diagonalizable** and $\mathbb{P}$ has N linearly independent eigenvectors, then $\mathbf{p}^{(n)}$ (distribution of X_n) will converge to this **unique stationary distribution** as time $n \to \infty$, regardless of the initial distribution.

$\lambda_1, \ldots, \underbrace{\lambda_p, \ldots \lambda_p}$

$AM(\lambda_p) = GM(\lambda_p)$

Power Iteration: Convergence

To converge to the eigenvector corresponding to the dominant eigenvalue of $\mathbb{P}$.

Assumptions:

1. The eigenvalues of the matrix $\mathbb{P}$ satisfies: (using P-F Th.)

$$\lambda_1 = 1 > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_N|.$$

   Let, $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_N$ be the corresponding left eigenvectors of $\mathbb{P}$.

2. Initial distribution $\underline{p}^{(0)}$ has a non-zero component in the direction of the eigenvector along the dominant eigenvalue, i.e.

$$\underline{p}^{(0)} = \theta_1 \underline{v}_1 + \theta_2 \underline{v}_2 + \ldots + \theta_N \underline{v}_N \text{ with } \theta_1 \neq 0.$$

$$\begin{aligned}
\underline{p}^{(n)} &= \underline{p}^{(n-1)}\mathbb{P} \\
&= \underline{p}^{(n-2)}\mathbb{P}^2 \\
&= \cdots \\
&= \underline{p}^{(0)}\mathbb{P}^n \\
&= \sum_{i=1}^{n} \theta_i \underline{v}_i \mathbb{P}^n = \sum_{i=1}^{n} \theta_i \lambda_i^n \underline{v}_i
\end{aligned}$$

$$\underline{v}_i \mathbb{P} = \lambda_i \underline{v}_i$$

$$\underline{v}_i \mathbb{P}^2 = \lambda_i \underline{v}_i \mathbb{P} = \lambda_i^2 \underline{v}_i$$

$$\underline{v}_i \mathbb{P}^n = \lambda_i^n \underline{v}_i$$

Using Condition 1, $\lim_{n\to\infty} \underline{p}^{(n)} = \lim_{n\to\infty} \sum_{i=1}^{n} \theta_i \lambda_i^n = \theta_1 \underline{v}_1$

$$\pi \mathbb{P} = \theta_1 \underline{v}_1 \mathbb{P} = \theta_1 \underline{v}_1 = \pi$$

Thus

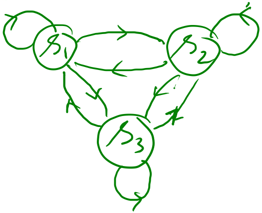
$$\boxed{\pi = \theta_1 \underline{v}_1} \text{ satisfies } \pi\mathbb{P} = \pi.$$

Note: θ_1 is the normalizing factor for π to be a p.m.f.

Example: Stationary Distribution

Ex 1: Consider the MC with the transition matrix:

$$\mathbb{P} = \begin{array}{cc} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix} \end{array}.$$



Show that the stationary distribution of the MC is

$$\underline{\pi} = \left[\frac{21}{62}, \frac{23}{62}, \frac{18}{62}\right] \approx [0.3387, 0.3710, 0.2903].$$

Initialize : $\underline{p}^{(0)} = [1/3, 1/3, 1/3]$

Iteration :

$$\underline{p}^{(1)} = \underline{p}^{(0)}\mathbb{P} = [0.3333, 0.3667, 0.3000]$$
$$\underline{p}^{(2)} = \underline{p}^{(1)}\mathbb{P} = [0.3367, 0.3700, 0.2933]$$
$$\underline{p}^{(3)} = \underline{p}^{(2)}\mathbb{P} = [0.3380, 0.3707, 0.2913]$$
$$\underline{p}^{(4)} = \underline{p}^{(3)}\mathbb{P} = [0.3385, 0.3709, 0.2907]$$
$$\underline{p}^{(5)} = \underline{p}^{(4)}\mathbb{P} = [0.3386, 0.3709, 0.2904]$$
$$\underline{p}^{(6)} = \underline{p}^{(5)}\mathbb{P} = [0.3387, 0.3710, 0.2904]$$
$$\underline{p}^{(7)} = \underline{p}^{(6)}\mathbb{P} = [0.3387, 0.3710, 0.2903] = \underline{\pi}$$

Stopping Criterion: $\boxed{\|\underline{p}^{(i+1)} - \underline{p}^{(i)}\|_1 < \varepsilon}$ — 0.000001