
Markov chain and transition matrix

1 Markov chain

Definition:

Let T be a $k \times k$ matrix with elements $\{T_{ij} : i, j = 1, \dots, n\}$. A random process (X_1, X_2, \dots) with finite state space $S = \{s_1, \dots, s_n\}$ is said to be a homogeneous Markov chain with transition matrix T , if for all k , all $i, j \in \{1, \dots, n\}$ and all $i_0, \dots, i_{k-1} \in \{1, \dots, n\}$ we have:

$$P(X_{k+1} = s_j | X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{k-1} = s_{i_{k-1}}, X_k = s_i) = P(X_{k+1} = s_j | X_k = s_i) = T_{ij}.$$

The elements of the transition matrix T are called transition probabilities.

Every transition matrix satisfies:

(1) $T_{ij} \geq 0$ for all $i, j \in \{1, \dots, n\}$, and

(2) $\sum_{j=1}^n T_{ij} = 1$ for all $i \in \{1, \dots, n\}$

There is another important characteristic of a Markov chain, namely the initial distribution, which tells us how the Markov chain starts. The initial distribution is represented as a row vector:

$$\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_n^{(0)}) = (P(X_0 = s_1), P(X_0 = s_2), \dots, P(X_0 = s_n))$$

and we have:

$$\sum_{i=1}^n \mu_i^{(0)} = 1$$

Theorem 1. For a Markov chain (X_0, X_1, \dots) with state space $S = \{s_1, \dots, s_n\}$, initial distribution $\mu^{(0)}$ and transition matrix T , we have for any k that the distribution $\mu^{(k)}$ at time n satisfies:

$$\mu^{(k)} = \mu^{(0)} T^k.$$

Now let us assume momentarily that for a given homogeneous Markov Chain with transition matrix T and initial probability distribution μ_0 there exists a limit distribution $\pi \in [0, 1]^n$ such that $\lim_{t \rightarrow \infty} \mu^{(t)} = \pi$

Then it must be the case that:

$$\pi = \lim_{t \rightarrow \infty} \mu^{(0)} T^t = \lim_{t \rightarrow \infty} \mu^{(0)} T^{t+1} = (\lim_{t \rightarrow \infty} \mu^{(0)} T^t) T = \pi T$$

Thus, any limit distribution is a left eigenvector of the transition matrix with eigenvalue 1, and can be computed by solving the equation $\pi = \pi T$. Solutions to this equation are called the equilibrium or stationary distributions of the chain.

2 Second largest eigenvalue

Theorem 2. Let x be a vector of length n such that $x_i = 1$ for all $i = 1, \dots, n$. Then $(Tx)_i = \sum_{j=1}^n a_{ij} x_j = \sum_{j=1}^n a_{ij} 1 = 1 = x_i$. Consequently $\lambda = 1$ is an eigenvalue.

Then we'll prove that $\lambda = 1$ is the largest eigenvalue for transition matrix T .

Theorem 3. If λ is an eigenvalue of a stochastic matrix T then $\lambda \leq 1$.

Proof. Let x be a right eigenvector corresponding to eigenvalue λ and let $x_k = \max_{i \in n} x_i$. Since $Tx = \lambda x$, so $(\lambda x)_k = (Tx)_k = p_{k1}x_1 + \dots + p_{kn}x_n$. It follows that

$$|\lambda| |x_k| = |p_{k1}x_1 + \dots + p_{kn}x_n| \leq |p_{k1}x_1| + \dots + |p_{kn}x_n| = p_{k1}|x_1| + \dots + p_{kn}|x_n| \leq |x_k|(p_{k1} + \dots + p_{kn}) = |x_k| \text{ which proves the claim. } \square$$

Convergence of Regular Markov Chains

Theorem 4. (Jordan canonical form)

Let $A \in C^{n \times n}$ be any matrix with real and distinct eigenvalues $\lambda_1, \dots, \lambda_n \in R$, $l \leq n$. Then there

$$\text{exists an invertible matrix } U \in C^{n \times n} \text{ such that } UAU^{-1} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

In this case one notes that in fact the columns of $U^{-1} = V$ are precisely the right eigenvectors corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$. It has a unique largest eigenvalue $\lambda_1 = 1$, and the other eigenvalues may be ordered so that $1 \geq |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$. The unique left eigenvector associated to eigenvalue 1 is the stationary distribution π , and the corresponding unique right eigenvector is $\mathbf{1} = (1, 1, \dots, 1)$. If the first row of U is normalised to π , then the first column of V must be normalised to $\mathbf{1}$ because $UV = UU^{-1} = I$, and hence $(UV)_{11} = u_1 v_1 = \pi v_1 = 1$.

Denoting

$$\Lambda = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

We have then:

$$P^2 = (V\Lambda U)^2 = V\Lambda^2 U = V \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \lambda_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^2 \end{bmatrix} U$$

$$\text{And in general } P^t = (V\Lambda U)^t = V\Lambda^t U = V \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \lambda_2^t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^t \end{bmatrix} U \rightarrow U = \pi \text{ as } t \rightarrow \infty$$

To make the situation even more transparent, represent a given initial distribution $q = q_0$ in the (left) eigenvector basis as $q = \tilde{q}_1 \mu_1 + \tilde{q}_2 \mu_2 + \dots + \tilde{q}_n \mu_n = \pi + \tilde{q}_2 \mu_2 + \dots + \tilde{q}_n \mu_n$, where $\tilde{q}_i = \langle q^T, v_i \rangle = q^T v_i$

$$\text{Then } qP = (\pi + \tilde{q}_2 \mu_2 + \dots + \tilde{q}_n \mu_n)T = \pi + \tilde{q}_2 \lambda_2 \mu_2 + \dots + \tilde{q}_n \lambda_n \mu_n$$

$$\text{and generally } q^{(t)} = qT^t = \pi + \sum_{i=2}^n \tilde{q}_i \lambda_i^t \mu_i$$

implying that $q^{(t)} \rightarrow \pi$, and if the eigenvalues are ordered as assumed, then

$$\|q^{(t)} - \pi\| = o(|\lambda_2|^t).$$

By this we see that smaller second largest eigenvalue gives a higher rate of convergence.

3 Compute transition matrix for Markov chain

(1) Scalar input:

Input: observations- x , number of hidden states- n

Output: transition matrix- T

Initialzie: $T=0_{n \times n}$

Recursion: for t in range(len(x)): $p[x[t-1]-1, x[t]-1] = p[x[t-1]-1, x[t]-1] + 1$

(2) Vector input:

Input: observations-x(a matrix with each row is a one-hot vector), number of hidden states-n(length of one-hot vector)

Output: transition matrix-T

Initialzie: $T=0_{n \times n}$

Recursion: for t in range(len(x)): $p=p+np.outer(x[t-1],x[t])$

4 Stochastic matrix

A right stochastic matrix is a real square matrix, with each row summing to 1.

A left stochastic matrix is a real square matrix, with each column summing to 1.

A doubly stochastic matrix is a square matrix of nonnegative real numbers with each row and column summing to 1.

5 Generate doubly stochastic matrix

Intuition:

Recall that X is doubly stochastic matrix if and only if $Xe_n = e_n$ and $e_n^T X = e_n^T$ if and only if $XJ_n = J_n X = J_n$, where J_n is the $n \times n$ matrix with each of its entries is equal to $1/n$. Hence for any doubly stochastic matrix X , e_n is an eigenvector for both X and X^T corresponding to the eigenvalue 1.

for any V such that the first row of V is equal to a multiple of e_n^T , and each of the last $n - 1$ row sums is zero, then clearly, its inverse V^{-1} has its first column equal to a multiple of e_n and each of the last $n - 1$ column sums of V^{-1} is zero. Such V is said to have pattern S .

Theorem 5. For transition matrix T , get the eigenvalues: $1, \lambda_2, \dots, \lambda_k$

If $1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq -1$

and $1 - (n - 1)\lambda_2 + \lambda_3 + \dots + \lambda_n \geq 0$,

$1 + (n - 1)\lambda_n \geq 0$,

then there is an $n \times n$ nonsymmetric doubly stochastic matrix D with spectrum $1, \lambda_2, \dots, \lambda_k$.

$$\text{For matrix } V = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

$$\text{and its inverse } V^{-1} \text{ is given by: } V^{-1} = \begin{bmatrix} 1/n & 1/n & 1/n & \dots & 1/n \\ 1/n & -(n-1)/n & 1/n & \dots & 1/n \\ 1/n & 1/n & -(n-1)/n & \dots & 1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/n & 1/n & 1/n & \dots & -(n-1)/n \end{bmatrix}$$

Now the entries of the matrix $A = a_{ij} = V\Lambda V^{-1}$ satisfy the following relations:

$$a_{11} = 1/n(\text{trace}(\Lambda))$$

$$a_{ii} = 1/n(1 + (n - 1)\lambda_i) \text{ for } i = 2, \dots, n$$

$$a_{i1} = 1/n(1 + \lambda_2 + \dots + \lambda_{i-1} - (n - 1)\lambda_i + \lambda_{i+1} + \dots + \lambda_n)$$

$$a_{ij} = 1/n(1 - \lambda_j) \text{ for } j \geq 2 \text{ and } j \neq i$$

6 Create synthetic data for different eigenvalue

According to the Theorem4, we can conclude that:

$$Q * \Lambda * Q^{-1} \text{ has the same eigenvalues as } \Lambda, \text{ and } \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

So for different eigenvalues Λ , we can create synthetic data by randomly generate a orthonormal matrix Q and $Q * \Lambda * Q^{-1}$ is the synthetic data.

References

- [1] Mourad, Bassam, et al. "An algorithm for constructing doubly stochastic matrices for the inverse eigenvalue problem." Linear Algebra and its Applications 439.5 (2013): 1382-1400.
- [2] Combinatorial Models and Stochastic Algorithms Lecture Notes. Pekka Orponen Helsinki University of Technology, Laboratory for Theoretical Computer Science