Markov chain and transition matrix

1 Markov chain

Definition:

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

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

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

Every transition matrix satisfies:

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

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

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:

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

and we have:

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

Theorem 1. *For a Markow chain $(X_0, X_1, ...)$ with state space $S = \{s_1, ..., s_n\}$, initial distribution $\mu^{(0)}$ and transition matirx T, we have for any k that the distribution $\mu^{(n)}$ 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 $limit_{t \to \infty} \mu^{(t)} = \pi$

Then it must be the case that:

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

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 $i = 1, ..., n$. 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 corresonding to eigenvalue λ and let $x_k = max_{i \in n} x_i$. Since $Tx = \lambda x$, so $(\lambda x)_k = (Tx)_k = p_{k1}x_1 + ... + p_{kn}x_n$ It follows that

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

Theorem 4. *(Spectral Decomposition)*

Let M be a real symmetric R^{d*d} matrix with eigenvalues $\lambda_1, ..., \lambda_d$ and corresponding orthonormal eigenvectors $u_1, ..., u_d$, and $Q = [u_1, ..., u_d]$, $\Lambda = diag(\lambda_1, ..., \lambda_d)$. Then:

$M = Q\Lambda Q^T$ *and* $M = \sum_{i=1}^{d} \lambda_i \mu_i \mu_i^T$

Proof. $Q\Lambda Q^T \mu_i = Q\Lambda e_i = Q\lambda_i e_i = \lambda_i \mu_i = M\mu_i$. Thus $Q\Lambda Q^T = M$

And for any j, $(\sum_i \lambda_i \mu_i \mu_i^T)\mu_j = \lambda_j \mu_j = M\mu_j$

Hence $M = Q\Lambda Q^T$ and $M = \sum_{i=1}^{d} \lambda_i \mu_i \mu_i^T$. □

Convergence of Regular Markov Chains

Theorem 5. *(Jordan canonical form)*

Let $A \in C^{n*n}$ be any matrix with eigenvalues$\lambda_1, ..., \lambda_l \in C$, $l \leq n$. Then there exists an invertible matrix $U \in^{n*n}$ such that $UAU^{-1} = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_l \end{bmatrix}$

where each J_i is a $k_i * k_i$ Jordan block associated to some eigenvalue λ of A: $J_i = \begin{bmatrix} \lambda & 1 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda \end{bmatrix}$

A regular Markov chain with transition matrix T has a unique stationary distribution vector π such that $\pi T = \pi$. Assume for simplicity that all the eigenvalues of T are real and distinct. Then the rows of U may be taken to be left eigenvectors of the matrix P, and the Jordan canonical form reduces to the familiar eigenvalue decomposition:

$$UPU^{-1} = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

In this case one notes that in fact the columns of $U^1 = V$ are precisely the right eigenvectors corresponding to the eigenvalues $\lambda_1, ..., \lambda_n$. T has a unique largest eigenvalue $\lambda_1 = 1$, and the other eigenvalues may be ordered so that $1 \geq |\lambda_2| \geq |\lambda_3| \geq ... \geq |\lambda_n|$. The unique left eigenvector associated to eigenvalue 1 is the stationary distribution π, and the corresponding unique right eigenvector is $1 = (1, 1, ..., 1)$. If the first row of U is normalised to π, then the first column of V must be normalised to $\vec{1}$ because $UV = UU^1 = I$, and hence $(UV)_{11} = u_1 v_1 = \pi v1 = 1$.

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:

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

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

To make the situation even more transparent, represent a given initial distribution $q = q_0$ in the (left) eigenvector basis as $q = \widetilde{q_1}\mu_1 + \widetilde{q_2}\mu_2 + ... + \widetilde{q_n}\mu_n = \pi + \widetilde{q_2}\mu_2 + ... + \widetilde{q_n}\mu_n$, where $\widetilde{q_i} =< q^T, v_i >= qv_i$

Then $qP = (\pi + \widetilde{q_2}\mu_2 + ... + \widetilde{q_n}\mu_n)T = \pi + \widetilde{q_2}\lambda_2\mu_2 + ... + \widetilde{q_n}\lambda_n\mu_n$

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

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

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

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*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*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

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

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

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

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

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

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

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

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

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

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

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

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

6 Create synthetic data for different eigenvalue

According to the spectral decompostion, we can conclude that:

if Q is orthogonal, then $Q * \Lambda * Q^T$ has the same eigenvalues as A, 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 $Q * \Lambda * Q^T$ 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