Markov Chains

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

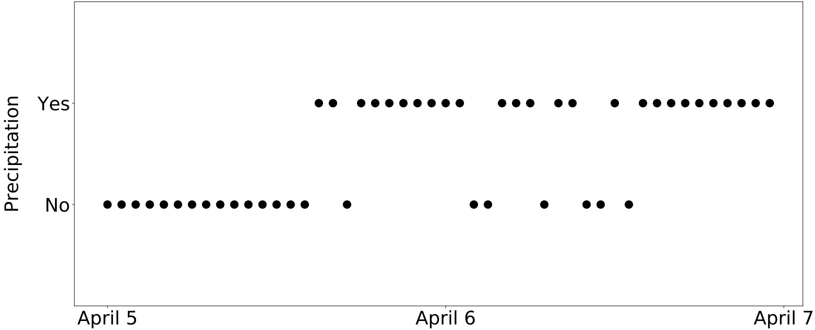
Goal

Model time series

Data: $x_1, x_2, \ldots, x_n$

x_i is measurement at time i

Precipitation data



First idea

Represent precipitation at each time by a random variable $\tilde{a}_i$

Then estimate joint pmf of $\tilde{a}_1, \ldots, \tilde{a}_n$ from data

Entries in joint pmf? 2^n (if $n = 100$ more than 10^{30}!)

Curse of dimensionality

Second idea

Assume data are i.i.d. random variables with pmf $p_{\tilde{a}}$

$$p_{\tilde{a}_1,\ldots,\tilde{a}_n} = \prod_{i=1}^{n} p_{\tilde{a}}$$

Good news: We only need to estimate one parameter

Bad news: We aren't modeling temporal structure

Third idea

Just model transitions

Markov property

$\tilde{a}_1$, $\tilde{a}_2$, ..., $\tilde{a}_n$ satisfy the Markov property if:

$\tilde{a}_{i+1}$ is conditionally independent of $\tilde{a}_1, \ldots, \tilde{a}_{i-1}$ given $\tilde{a}_i$

$$p_{\tilde{a}_{i+1} \,|\, \tilde{a}_1, \ldots, \tilde{a}_i} \left(a_{i+1} \,|\, a_1, a_2, \ldots, a_i \right) = p_{\tilde{a}_{i+1} \,|\, \tilde{a}_i} \left(a_{i+1} \,|\, a_i \right)$$

$$\begin{aligned} & p_{\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n} \left(a_1, a_2, \ldots, a_n \right) \\ & = p_{\tilde{a}_1} \left(a_1 \right) p_{\tilde{a}_2 \,|\, \tilde{a}_1} \left(a_2 \,|\, a_1 \right) p_{\tilde{a}_3 \,|\, \tilde{a}_1, \tilde{a}_2} \left(a_3 \,|\, a_1, a_2 \right) \\ & \quad p_{\tilde{a}_4 \,|\, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3} \left(a_4 \,|\, a_1, a_2, a_3 \right) \ldots \\ & = p_{\tilde{a}_1} \left(a_1 \right) p_{\tilde{a}_2 \,|\, \tilde{a}_1} \left(a_2 \,|\, a_1 \right) p_{\tilde{a}_3 \,|\, \tilde{a}_2} \left(a_3 \,|\, a_2 \right) \\ & \quad p_{\tilde{a}_4 \,|\, \tilde{a}_3} \left(a_4 \,|\, a_3 \right) \ldots \end{aligned}$$

Third idea

Markov chain model

Requires estimating $p_{\tilde{a}_1}$ and $p_{\tilde{a}_{i+1} \mid \tilde{a}_i}(a_{i+1} \mid a_i)$ for $1 \leq i \leq n-1$

Good news: We only need to estimate $2n-1$ parameters

Bad news: We have n data points

Fourth idea

Assume transition probabilities are all the same

$$p_{\tilde{a}_{i+1}\,|\,\tilde{a}_i}\left(a_{i+1}\,|\,a_i\right) = p_{\text{cond}}\left(a_{i+1}\,|\,a_i\right) \qquad 1 \leq i \leq n-1$$

$$\begin{aligned}
&p_{\tilde{a}_1,\tilde{a}_2,\ldots,\tilde{a}_n}\left(a_1, a_2, \ldots, a_n\right) \\
&= p_{\tilde{a}_1}\left(a_1\right) p_{\tilde{a}_2\,|\,\tilde{a}_1}\left(a_2\,|\,a_1\right) p_{\tilde{a}_3\,|\,\tilde{a}_2}\left(a_3\,|\,a_2\right) p_{\tilde{a}_4\,|\,\tilde{a}_3}\left(a_4\,|\,a_3\right)\ldots \\
&= p_{\tilde{a}_1}\left(a_1\right) p_{\text{cond}}\left(a_2\,|\,a_1\right) p_{\text{cond}}\left(a_3\,|\,a_2\right) p_{\text{cond}}\left(a_4\,|\,a_3\right)\ldots
\end{aligned}$$

Number of parameters: 3

Empirical probabilities (2015)

Marginal probabilities

No	88.7
Yes	11.3

1-step conditional probabilities

		Hour h	
		No	**Yes**
Hour $h+1$	**No**	96.0	31.2
	Yes	4.0	68.8

2-step conditional probabilities

No precipitation at $h-1$

		Hour h	
		No	**Yes**
Hour $h+1$	**No**	97.1	49.4
	Yes	2.9	50.6

Precipitation at $h-1$

		Hour h	
		No	**Yes**
Hour $h+1$	**No**	70.6	23.0
	Yes	29.4	77.0

Task: Predict precipitation in 2016

Marginal probabilities

No	88.7
Yes	11.3

Test accuracy (2016): 83.4%

1-step conditional probabilities

		Hour h	
		No	**Yes**
Hour $h+1$	**No**	96.0	31.2
	Yes	4.0	68.8

Test accuracy (2016): 87.3%

2-step conditional probabilities

No precipitation at $h-1$

		Hour h	
		No	**Yes**
Hour $h+1$	**No**	97.1	49.4
	Yes	2.9	50.6

Precipitation at $h-1$

		Hour h	
		No	**Yes**
Hour $h+1$	**No**	70.6	23.0
	Yes	29.4	77.0

Test accuracy (2016): 87.3%

Finite state Markov chain

Each entry takes value in finite set of *states* $\{s_1, \ldots, s_m\}$

Marginal pmf represented by state vector:

$$\pi_i := \begin{bmatrix} p_{\tilde{a}_i}(s_1) \\ p_{\tilde{a}_i}(s_2) \\ \cdots \\ p_{\tilde{a}_i}(s_m) \end{bmatrix}.$$

Time homogeneous finite state Markov chain

All transition probabilities are the same

$$p_{\tilde{a}_{i+1} \,|\, \tilde{a}_i}\left(a_{i+1} \,|\, a_i\right) = p_{\text{cond}}\left(a_{i+1} \,|\, a_i\right) \qquad 1 \leq i \leq n-1$$

Transition matrix

$$T := \begin{bmatrix} p_{\text{cond}}(s_1 \,|\, s_1) & p_{\text{cond}}(s_1 \,|\, s_2) & \cdots & p_{\text{cond}}(s_1 \,|\, s_m) \\ p_{\text{cond}}(s_2 \,|\, s_1) & p_{\text{cond}}(s_2 \,|\, s_2) & \cdots & p_{\text{cond}}(s_1 \,|\, s_m) \\ \cdots & \cdots & \cdots & \cdots \\ p_{\text{cond}}(s_m \,|\, s_1) & p_{\text{cond}}(s_m \,|\, s_2) & \cdots & p_{\text{cond}}(s_m \,|\, s_m) \end{bmatrix}$$

Precipitation

1-step conditional probabilities

		Hour h	
		No	**Yes**
Hour $h+1$	**No**	96.0	31.2
	Yes	4.0	68.8

Transition matrix

$$T := \begin{bmatrix} 0.960 & 0.312 \\ 0.040 & 0.688 \end{bmatrix}$$

Car rental

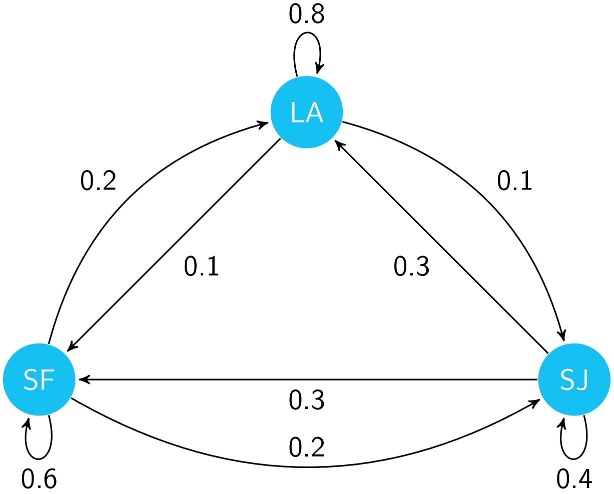
Goal: Model location of cars

3 locations (states): Los Angeles, San Francisco, San Jose

Transition probabilities:

$$\begin{array}{cccc} \text{San Francisco} & \text{Los Angeles} & \text{San Jose} & \\ \left(\begin{array}{c} 0.6 \\ 0.2 \\ 0.2 \end{array}\right. & \begin{array}{c} 0.1 \\ 0.8 \\ 0.1 \end{array} & \left.\begin{array}{c} 0.3 \\ 0.3 \\ 0.4 \end{array}\right) & \begin{array}{c} \text{San Francisco} \\ \text{Los Angeles} \\ \text{San Jose} \end{array} \end{array}$$

Car rental

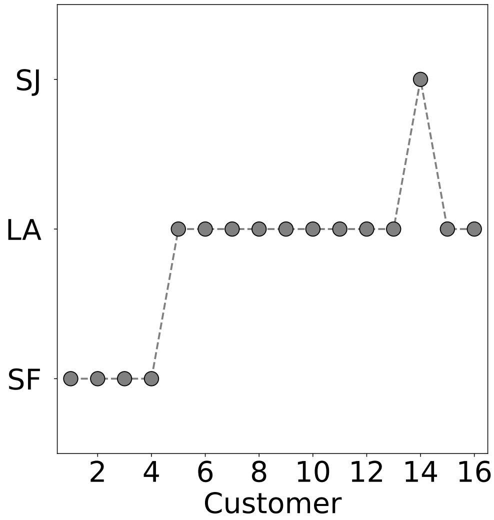


State vector and transition matrix

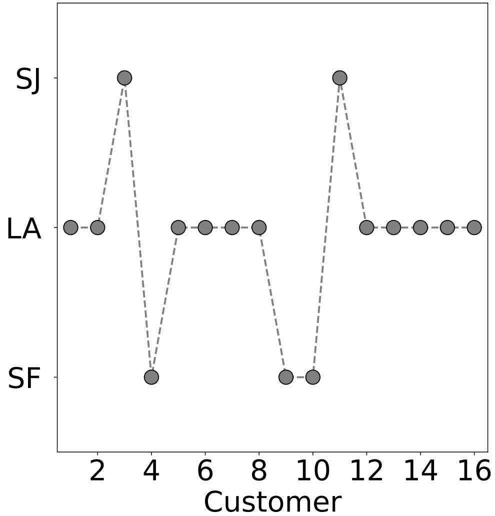
Cars are initially allocated to each location with same probability

$$\pi_1 := \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \qquad T := \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.2 & 0.1 & 0.4 \end{bmatrix}$$

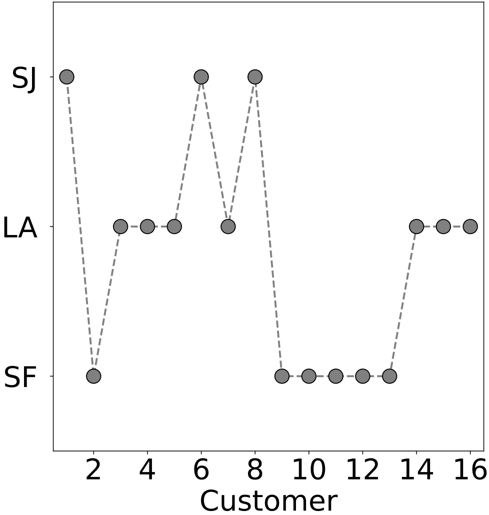
Realization



Realization



Realization



Computing probabilities

Probability that a car starts in SF and is in SJ after the 2nd customer

$$\begin{aligned} p_{\tilde{a}_1,\tilde{a}_3}(1,3) &= \sum_{i=1}^{3} p_{\tilde{a}_1,\tilde{a}_2,\tilde{a}_3}(1,i,3) \\ &= \sum_{i=1}^{3} p_{\tilde{a}_1}(1)\, p_{\tilde{a}_2 \mid \tilde{a}_1}(i \mid 1)\, p_{\tilde{a}_3 \mid \tilde{a}_2}(3 \mid i) \\ &= \pi_1[1] \sum_{i=1}^{3} T_{i1} T_{3i} \\ &= \frac{0.6 \cdot 0.2 + 0.2 \cdot 0.1 + 0.2 \cdot 0.4}{3} \approx 7.33\, 10^{-2} \end{aligned}$$

State vector and transition matrix

$$\pi_i := \begin{bmatrix} p_{\tilde{a}_i}(s_1) \\ p_{\tilde{a}_i}(s_2) \\ \cdots \\ p_{\tilde{a}_i}(s_m) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} p_{\tilde{a}_{i-1}}(s_j)\, p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_1 \mid s_j) \\ \sum_{j=1}^{m} p_{\tilde{a}_{i-1}}(s_j)\, p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_2 \mid s_j) \\ \cdots \\ \sum_{j=1}^{m} p_{\tilde{a}_{i-1}}(s_j)\, p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_m \mid s_j) \end{bmatrix}$$

$$= \begin{bmatrix} p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_1 \mid s_1) & p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_1 \mid s_2) & \cdots & p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_1 \mid s_m) \\ p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_2 \mid s_1) & p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_2 \mid s_2) & \cdots & p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_2 \mid s_m) \\ & & \cdots & \\ p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_m \mid s_1) & p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_m \mid s_2) & \cdots & p_{\tilde{a}_i \mid \tilde{a}_{i-1}}(s_m \mid s_m) \end{bmatrix} \begin{bmatrix} p_{\tilde{a}_{i-1}}(s_1) \\ p_{\tilde{a}_{i-1}}(s_2) \\ \cdots \\ p_{\tilde{a}_{i-1}}(s_m) \end{bmatrix}$$

$$= T\pi_{i-1}$$

State vector and transition matrix

$$\begin{aligned}\pi_i &= T\,\pi_{i-1}\\ &= T\,T\,\pi_{i-2}\\ &= T^{i-1}\,\pi_1\end{aligned}$$

5th customer

$$\pi_1 := \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \qquad T := \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.2 & 0.1 & 0.4 \end{bmatrix}$$

$$\pi_6 = T^5 \pi_1 = \begin{bmatrix} 0.281 \\ 0.534 \\ 0.185 \end{bmatrix}$$

What we have learned

Why we use Markov chains to model time series

Basic properties of Markov chains