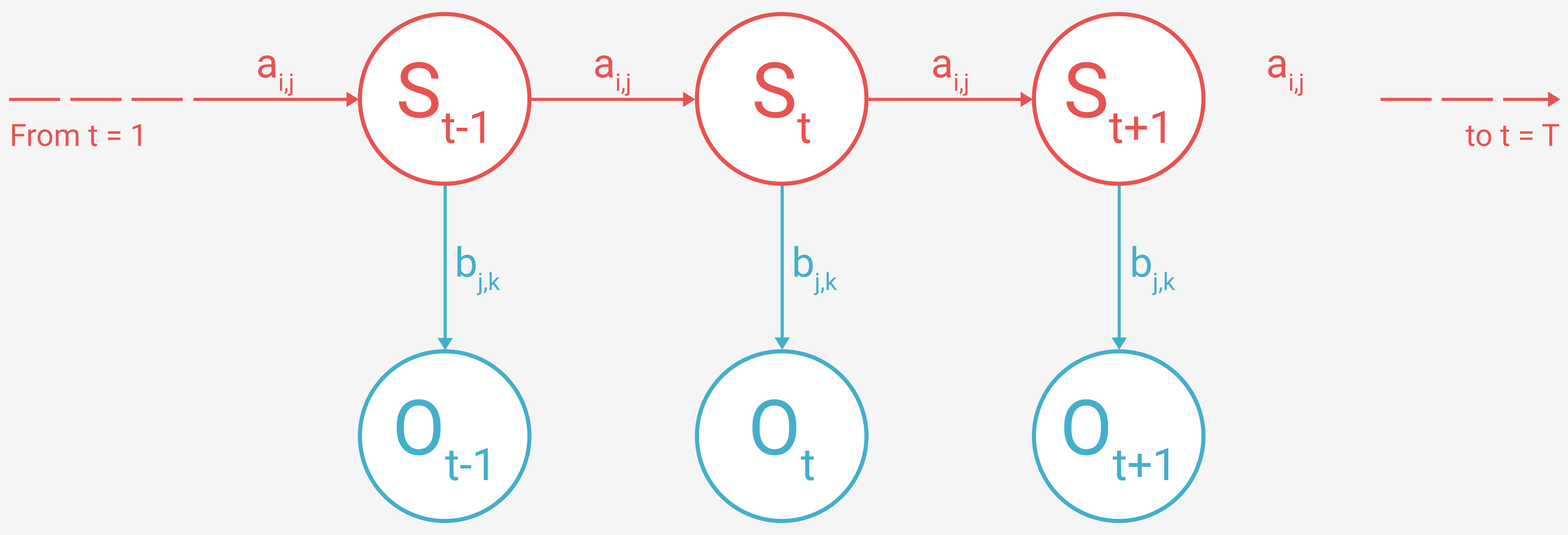
HMM - Terminology

General illustration of state-/ observation probabilities



Intuitive meaning

- If we are in state S_{t-1}, we use the *transition matrix* $a_{i,j}$ to find out what's the most probable state in the next time step t. If we know what state we are in, we can use the *emission matrix* $b_{j,k}$ to find out what's the most probable observation made in that same time step t-1.

Definitions

Detailed

- **Background info** The abbreviation **HMM** stands for **H**idden **M**arkov **M**odel. The word hidden is used because the states are hidden. We make observations that will help us calculate the probability of being in a certain state at a certain time step. There are other types of Markov Models as well, but these are not covered on this page.
- **Cardinality** When using the word cardinality we refer to the number of elements in a certain set. For example the cardinality of the set $\mathbf{S_1}$ = {**X**, **Y**, **Z**} is three because the set contains three unique elements. To be precise, the set $\mathbf{S_2}$ = {**X**, **Y**, **Z**, **Y**, **Z**, **X**} also have the cardinality three, because the set also contains three unique elements. In a mathematical context we write $|\ \mathbf{S_1}\ | = 3$ to denote that the cardinality of set $\mathbf{S_1}$ is three.
- **Time steps** When we work with HMM, time is discrete. That is, we work with discrete time steps where each timestep is an integer. The first time step is either declared as t_0 or t_1, depending on which source your reading. On this page you have the opportunity to decide if the content should be 0-indexed or 1-indexed (the button in the top-right corner). The last time step is declared as t_{T-1} if 0-indexing is used and t_T if using 1-indexing.
- **States** As mentioned above, states in HMM are hidden. That is, we can't for sure know what state is present in the current time step. Depending on which source you're reading, states are called different things. Some of them, if not all, are: *Hidden States*, *States* and *Emitters*. On this page, we will use the notation *States*. There are many different ways to mathematically notate a certain state. On this site, we declare a *state* at a certain time step *t* to be denoted S_t.
- **Observations** As with states, observations have different names depending on which source you're getting your information from. Some of them, if not all, are: *Outputs*, *Emissions*, *Observations* and *Visible States*. On this site, we use the word *observation*, and we denote an observation at time step *t* as O_t.
- **Set of time steps** The set of all time steps are defined as $\{t_0, t_1, \ldots, t_{T-1}\}$. Meaning that the cardinality of the set of time steps $|\ \{t_0, t_1, \ldots, t_{T-1}\}\ | = \mathbf{T}$.
- **Set of states** The set of all states are defined as $\{S_0, S_1, \ldots, S_{N-1}\}$. Meaning that the cardinality of the set of states $|\ \{S_0, S_1, \ldots, S_{N-1}\}\ | = \mathbf{N}$. On this page, we notate the set of all possible states as **S**.
- **Set of observations** The set of all observations are defined as $\{O_1, O_2, \ldots, O_T\}$. Meaning that the cardinality of the set of states $|\ \{O_0, O_1, \ldots, O_{K-1}\}\ | = \mathbf{K}$. On this page, we notate the set of all possible observations as **O**.
- **Row-stochastic** By using the word row-stochastic we mean that every row in the current matrix (or vector) must sum up to one. Let's define the following two matrices:

$$x_{i,j} = \begin{bmatrix} 0.30 & 0.35 & 0.53 \\ 0.43 & 0.12 & 0.31 \\ 0.12 & 0.48 & 0.37 \end{bmatrix} \qquad y_{i,j} = \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.25 & 0.25 & 0.50 \\ 0.10 & 0.10 & 0.80 \end{bmatrix}$$

Are these two matrices row-stochastic? Let's find out!

$$x_{0,0} + x_{0,1} + x_{0,2} = 0.30 + 0.35 + 0.53 = 1.18 \qquad y_{0,0} + y_{0,1} + y_{0,2} = 0.30 + 0.30 + 0.40 = 1.0$$
$$x_{1,0} + x_{1,1} + x_{1,2} = 0.43 + 0.12 + 0.31 = 0.86 \qquad y_{1,0} + y_{1,1} + y_{1,2} = 0.25 + 0.25 + 0.50 = 1.0$$
$$x_{2,0} + x_{2,1} + x_{2,2} = 0.12 + 0.48 + 0.37 = 0.97 \qquad y_{2,0} + y_{2,1} + y_{2,2} = 0.10 + 0.10 + 0.80 = 1.0$$

As we can see, the matrix $y_{i,j}$ is row-stochastic while $x_{i,j}$ is not.

- **Initial state distribution** The initial state distribution, which is sometimes called initial state probabilities, is a vector containing probabilities of being in a certain state at the initial time step t_0. The initial state distribution is denoted π and the vector is defined as $[\ \pi_0 \ \ \pi_1 \ \cdots \ \pi_{N-1}\ ]$, where **N** is the number of states. The initial state distribution vector is always row-stochastic. Let's make an example:

$$\pi(i) = \begin{bmatrix} 0.42 & 0.12 & 0.46 \end{bmatrix}$$

The initial state distribution given above should be interpreted in the following way: the probability of being in state S_0 at time t_0 is 0.42, the probability of being in state S_1 at time t_0 is 0.12 and the probability of being in state S_2 at time t_0 is 0.46. We can also see that the vector is row-stochastic because 0.42 + 0.12 + 0.46 = 1.0. If we want, we can also write the probabilities with the following notation:

$$\pi(S_0) = 0.42$$
$$\pi(S_1) = 0.12$$
$$\pi(S_2) = 0.46$$

- **Transition matrix** The
- **Observation matrix** The

Definitions

Brief