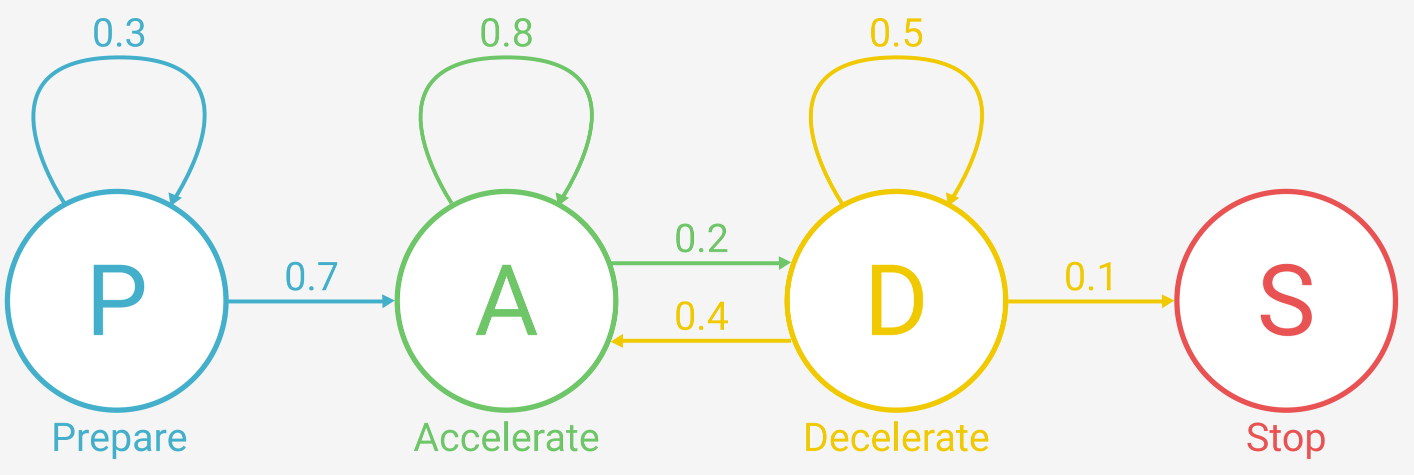
Runner example

← Back

Probabilities of state transitions



Intuitive meaning

- If Usain Bolt are preparing at t_x, the probability of him preparing at t_{x+1} is 0.3 and the probability of him accelerating at t_{x+1} is 0.7.
- If Usain Bolt are accelerating at t_x, the probability of him accelerating at t_{x+1} is 0.8 and the probability of him decelerating at t_{x+1} is 0.2.
- If Usain Bolt are decelerating at t_x, the probability of him decelerating at t_{x+1} is 0.5, the probability of him accelerating at t_{x+1} is 0.4, and the probability of him standing still at t_{x+1} is 0.1.
- If Usain Bolt are standing still at t_x, we can't make any more predictions.

Observation probabilities given state

Intuitive meaning

- If the HMM is in the prepare state, the probability of observing **observation 0** is 1.0.
- If the HMM is in the accelerate state, the probability of observing **observation 0** is 0.05, **observation 1** 0.38, **observation 2** 0.34, **observation 3** 0.20 and **observation 4** 0.03.
- If the HMM is in the decelerate state, the probability of observing **observation 1** is 0.05, **observation 2** 0.23, **observation 3** 0.28 and **observation 4** 0.44.
- If the HMM is in the stop state, the probability of observing **observation 4** is 0.18 and **observation 5** 0.82.

HMM parameters

- **Initial state distribution** As described in the *basic concepts* section, the initial state distribution declares the probabilities of being in a certain state at the initial time step t_0. In this example, the initial state will always be the prepare state, simply because it's reasonable to believe that Usain Bolt have to prepare before he starts running. Therefore the initial state distribution is given by:

$$\pi(i) = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

We define the set of states as S = {prepare, accelerate, decelerate, stop}, that is, S_0 = prepare, S_1 = accelerate, S_2 = decelerate and S_3 = stop. Now we have the information we need to express the initial state distribution in the following way:

$$\pi(\text{prepare}) = \pi(S_0) = P(S_0 = S_0) = 1.0$$
$$\pi(\text{accelerate}) = \pi(S_1) = P(S_0 = S_1) = 0.0$$
$$\pi(\text{decelerate}) = \pi(S_2) = P(S_0 = S_2) = 0.0$$
$$\pi(\text{stop}) = \pi(S_3) = P(S_0 = S_3) = 0.0$$

So what does this mean? In English we read the above notations as: *"at the initial time step t_0 the probability that Usain Bolt is preparing is 1.0, accelerating 0.0, decelerating 0.0, and standing still 0.0"*.

- **Transition matrix** We use the transition matrix A, or $a_{i,j}$, to describe the probabilities of transitioning between states. Note that it's possible for a HMM to transition from a certain state to itself. E.g. A[2][2] would be imply the probability $P(S_t = S_2 \mid S_{t-1} = S_2)$ which is a valid transition. In this example, we have the following transition matrix:

$$a_{i,j} = \begin{bmatrix} 0.3 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.2 & 0.0 \\ 0.0 & 0.4 & 0.5 & 0.1 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

Note that the last row is filled with zero. This implies that the last state is a **non-emitting** (also called **silent**) state. If the HMM reaches the non-emitting state, the model stops, because there is nowhere to go from that state. Also, note that when dealing with non-emitting states, we make an exception to the rule that the transition matrix need to be row-stochastic.

> We can't transition to any other state from a non-emitting/silent state.

To make the transition matrix a little clearer, let's have a look at the following figure which also describes the transition probabilities:

A	prepare	accelerate	decelerate	stop
prepare	0.3	0.7	0.0	0.0
accelerate	0.0	0.8	0.2	0.0
decelerate	0.0	0.4	0.5	0.1
stop	0.0	0.0	0.0	0.0

From this transition matrix we can see that $a_{1,2} = P(S_t = S_2 \mid S_{t-1} = S_1) =$
$= P(S_t =$ decelerate $\mid S_{t-1} =$ accelerate$) = 0.2$. Also, note that all rows, except the last, sum up to one!

- **Observation matrix** We use the observation matrix B, or $b_{j,k}$, to describe probabilities of making specific observations given that the hidden model is in a particular state. E.g. what is the probability that we observe given that the he are accelerating? In other words what's $P(O_t = O_1 \mid S_t = S_1) = P(O_t =$ $\mid S_t =$ accelerate$)$? The answer is 0.38. By looking at all the different states and their probabilities in relation to possible observations we get the observation matrix B:

$$b_{i,j} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.38 & 0.34 & 0.20 & 0.03 & 0.00 \\ 0.00 & 0.05 & 0.23 & 0.28 & 0.44 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.18 & 0.82 \end{bmatrix}$$

For better understanding we can have a look at the following figure, which we use to clearify the observation matrix:

B						
prepare	1.00	0.00	0.00	0.00	0.00	0.00
accelerate	0.05	0.38	0.34	0.20	0.03	0.00
decelerate	0.00	0.05	0.23	0.28	0.44	0.00
stop	0.00	0.00	0.00	0.00	0.18	0.82

From this observation matrix we can see that $b_{1,4} = P(O_t = O_4 \mid S_t = S_1) =$
$= P(O_t =$ $\mid S_t =$ accelerate$) = 0.03$. Also, note that all rows sum up to one!

- **Lambda** The initial state distribution π, the transition matrix A and the observation matrix B together describes how Usain Bolt transitions between the states and what observations we can expect when being in some of the states. At last, we can gather the three in $\lambda = \{A, B, \pi\}$.

Food for thought *Why does the observation matrix need to be row-stochastic?* To grasp this question we must first accept that each HMM is a simplification of reality. We make simplifications because it allows us to control all the variables in the system we are interested in. When an HMM are in a certain state, we *know* that the model will make some observation, it's *defined* that way. We have also *defined* that all possible observations that can be made are **O** = { , , , , , } . Therefore, in order for us to *know* that we make *some* observation, all separate possibilities for observations must together sum up to 1.0. The same reasoning can be made for the initial state distribution π and for the transition matrix A.