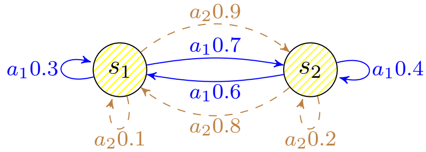
Discrete-time Markov Decision Process: 4-tuple $(\mathbb{S}, \mathbb{A}, \mathbf{T}, \mathbf{r})$

e.g.



$$\mathbb{S} = \{s_1, s_2\}, \mathbb{A} = \{a_1, a_2\}, \text{ Transition model } \mathbf{T}, \text{ Reward model } \mathbf{r}$$

Transition Model under Markovian Property: $T(s'|s, a)$, for discrete states we use transition matrix

$$\mathbf{T} = [T(s'|s,a)] = \begin{array}{cc} & \\ s_1 & \\ s_2 & \\ & \end{array} \begin{array}{c} \\ a_1 \\ a_2 \\ a_1 \\ a_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{bmatrix} 0.3 & 0.7 \\ 0.1 & 0.9 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{bmatrix} \end{array} = \begin{array}{c} \\ \\ a_1 \\ a_2 \\ \\ \end{array} \begin{array}{c} \\ s_1 \\ s_2 \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} o_1 & o_2 \end{array} \\ \begin{bmatrix} T_{111} & T_{112} \\ T_{121} & T_{122} \\ T_{211} & T_{212} \\ T_{221} & T_{222} \end{bmatrix} \end{array}$$

Reward Model: $r(s, a)$ or $r(s, a, s')$ below is an example of deterministic reward table:

$$\mathbf{r} = [r(s,a)] = \begin{array}{c} \\ \\ s_1 \\ s_2 \\ \\ \end{array} \begin{array}{c} \\ a_1 \\ a_2 \\ a_1 \\ a_2 \end{array} \begin{array}{c} r \\ \begin{bmatrix} r_{11} \\ r_{12} \\ r_{21} \\ r_{22} \end{bmatrix} \end{array} \text{ or } \mathbf{r} = [r(s,a,s')] = \begin{array}{c} \\ \\ s_1 \\ s_2 \\ \\ \end{array} \begin{array}{c} \\ a_1 \\ a_2 \\ a_1 \\ a_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{bmatrix} r_{111} & r_{112} \\ r_{121} & r_{122} \\ r_{211} & r_{212} \\ r_{221} & r_{222} \end{bmatrix} \end{array}$$

Sensor model: Conditional Probability of Observation fully observable

$$\mathbf{O} = [O(o'|s')] = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} o_1 & o_2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$$