Guided Tour of Machine Learning in Finance

Week 4: Reinforcement Learning

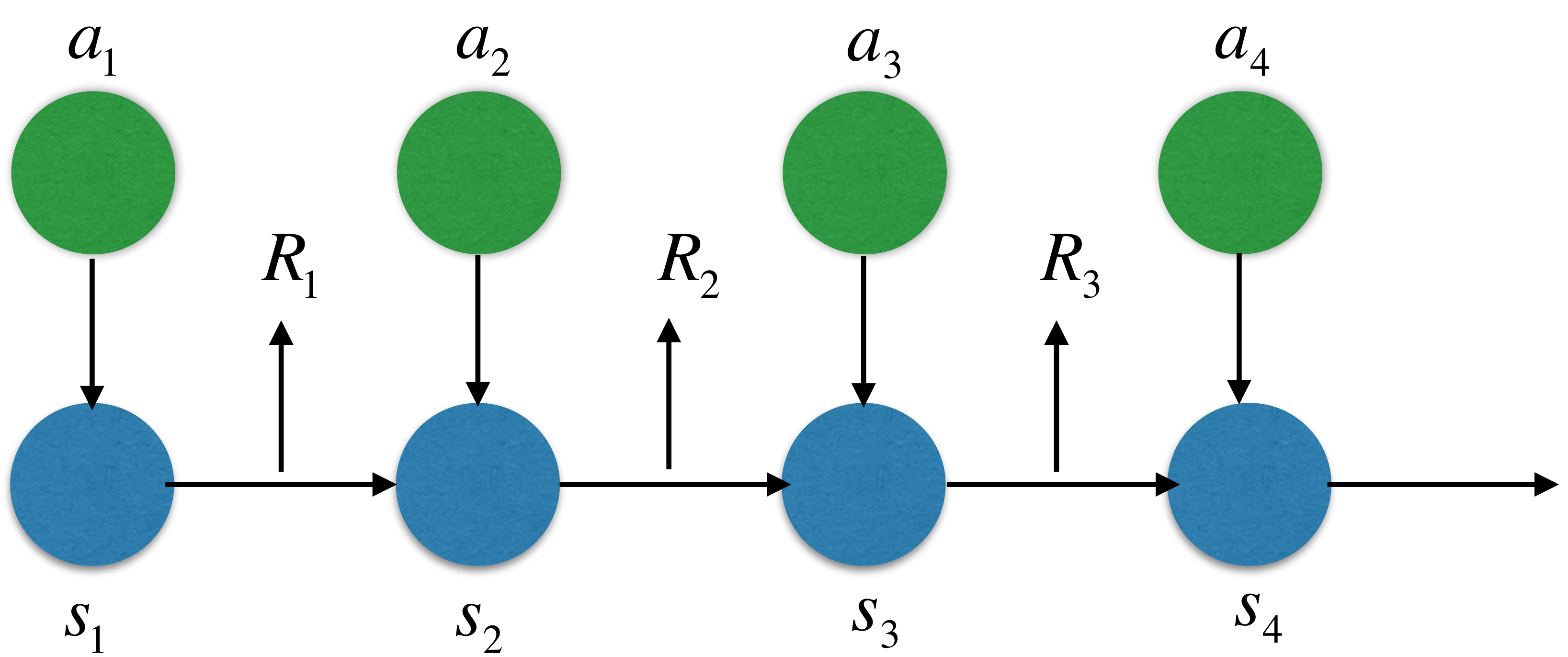
4-2-3-MDP-and-RL

Igor Halperin

NYU Tandon School of Engineering, 2017

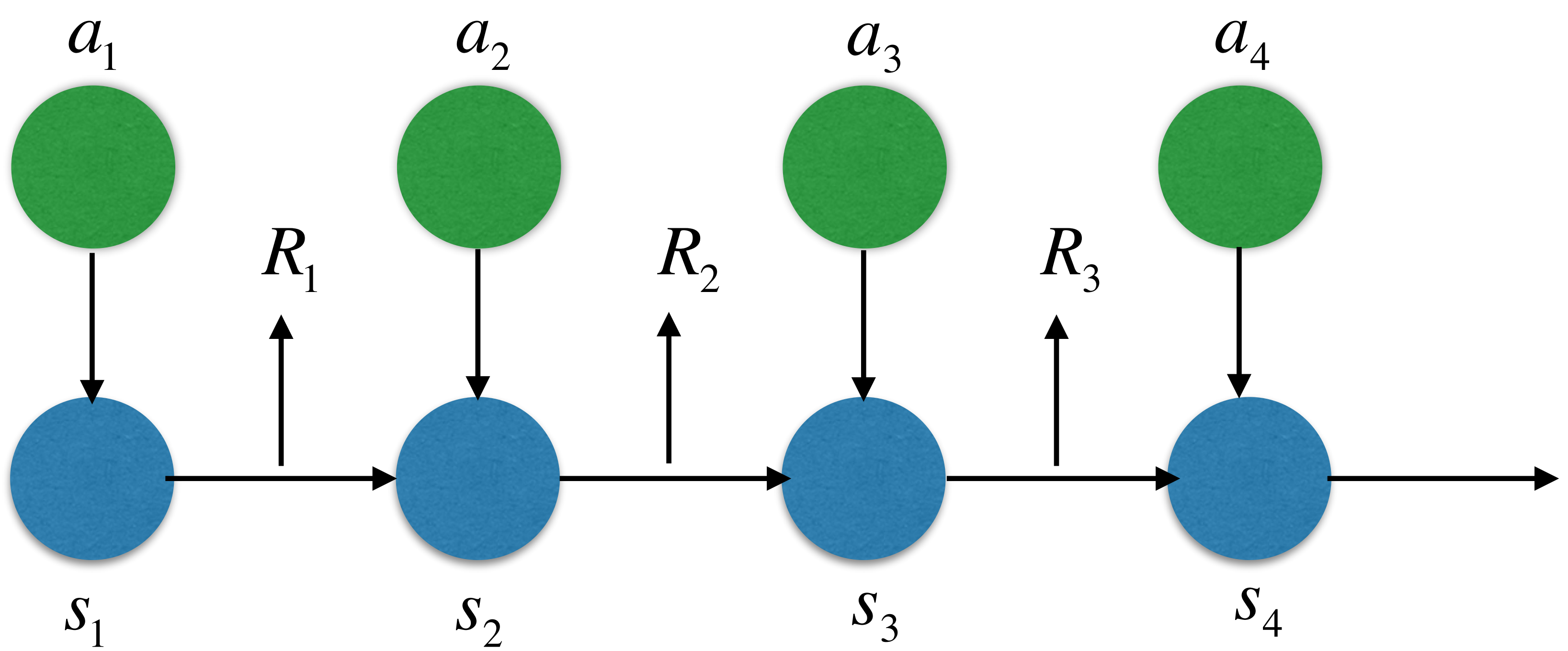
Markov Decision Processes

Markov Decision Processes: s_t - the observable environment whose dynamics can be modulated by agent's actions a_t



Markov Decision Processes

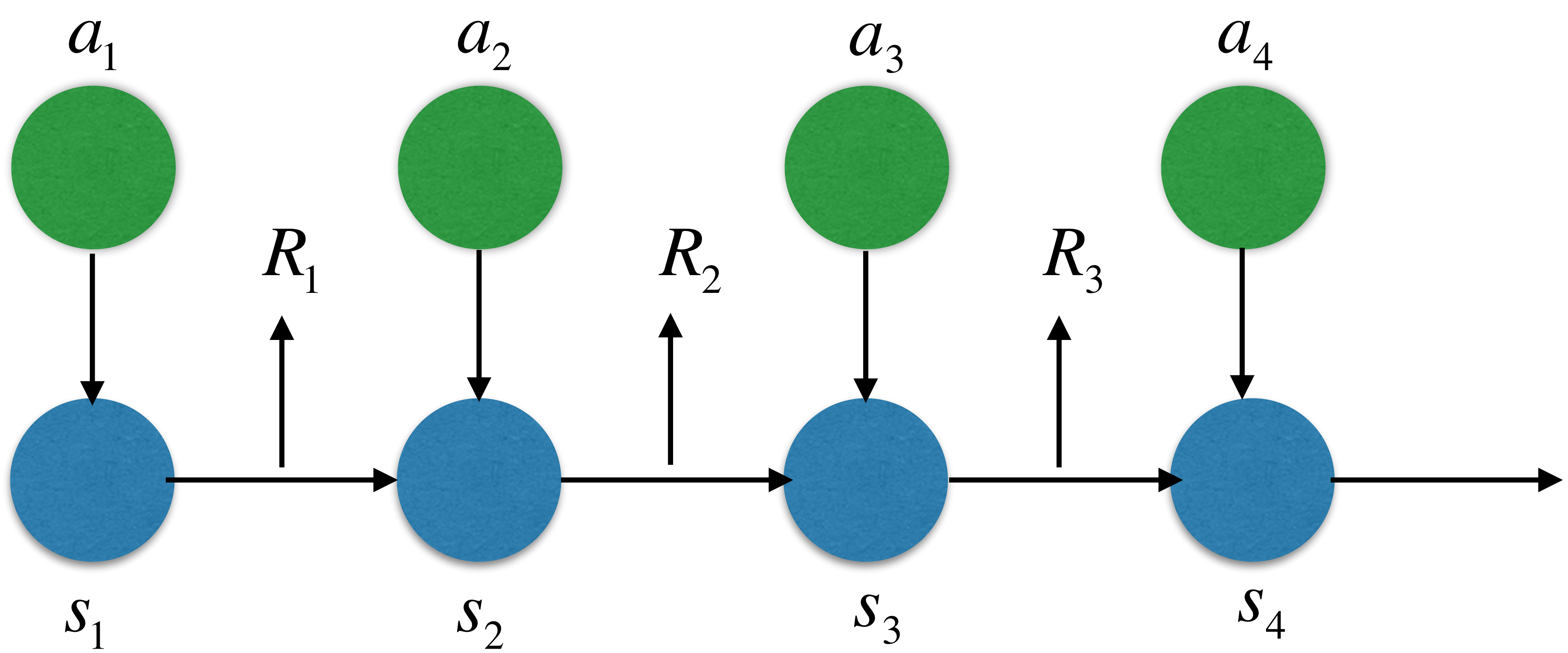
Markov Decision Processes: s_t - the observable environment whose dynamics can be modulated by agent's actions a_t



- $s_t \in S$: S is a set of **states** (discrete or continuous)
- $a_t \in A$: A is a set of **actions** (discrete or continuous)
- $p\left(s_{t+1} \mid s_t, a_t\right)$ are **transition probabilities**
- $R : S \times A \mapsto \mathbb{R}$ is a **reward function** (can depend on both state and action)
- $\gamma \in [0,1]$ is a **discount factor**

Markov Decision Processes

Markov Decision Processes: s_t - the observable environment whose dynamics can be modulated by agent's actions a_t

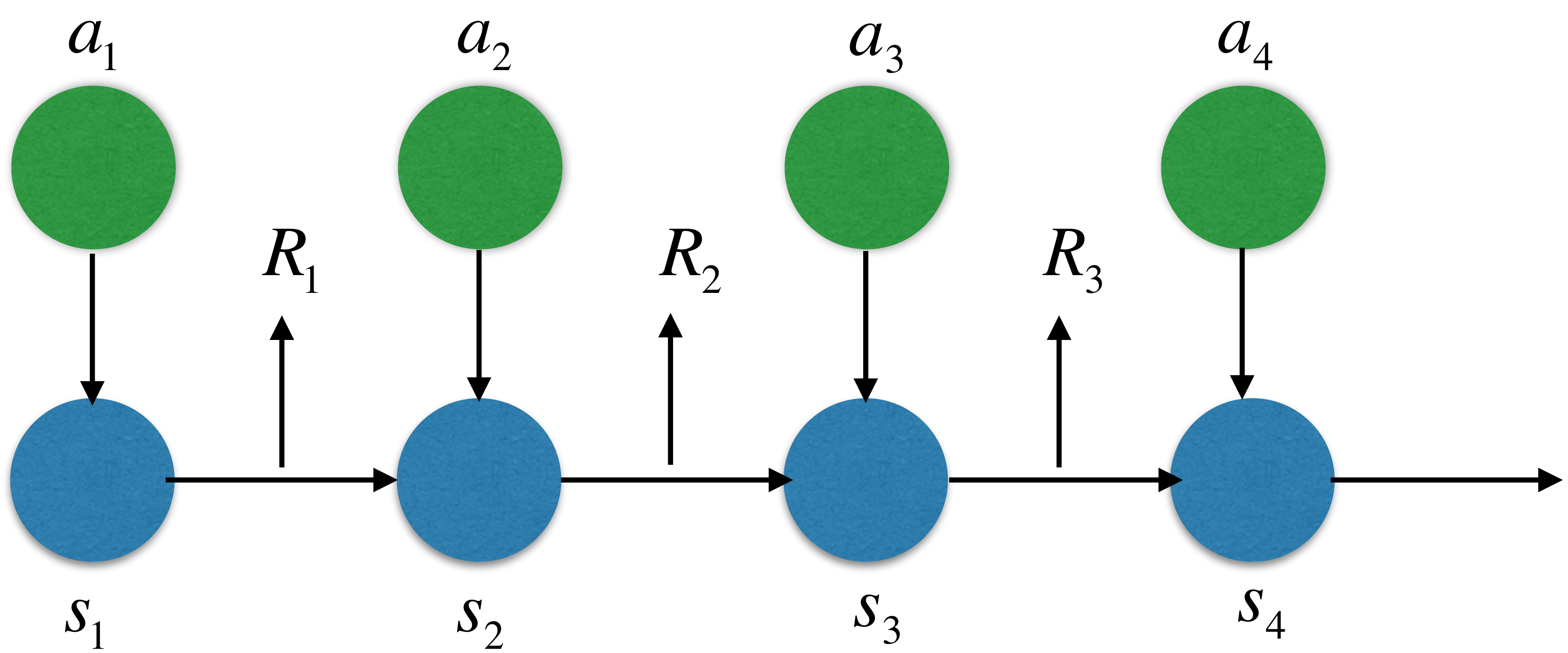


- $s_t \in S$: S is a set of **states** (discrete or continuous)
- $a_t \in A$: A is a set of **actions** (discrete or continuous)
- $p\left(s_{t+1} \mid s_t, a_t\right)$ are **transition probabilities**
- $R : S \times A \mapsto \mathbb{R}$ is a **reward function** (can depend on both state and action)
- $\gamma \in [0,1]$ is a **discount factor**
- **Cumulative total reward**

$$R(s_0, a_0) + \gamma R(s_1, a_1) + \gamma^2 R(s_2, a_2) + \ldots = \sum_n \gamma^n R(s_n, a_n)$$

RL: risk-neutral vs risk-sensitive

$$R(s_0,a_0)+\gamma R(s_1,a_1)+\gamma^2 R(s_2,a_2)+\ldots=\sum_n \gamma^n R(s_n,a_n)$$



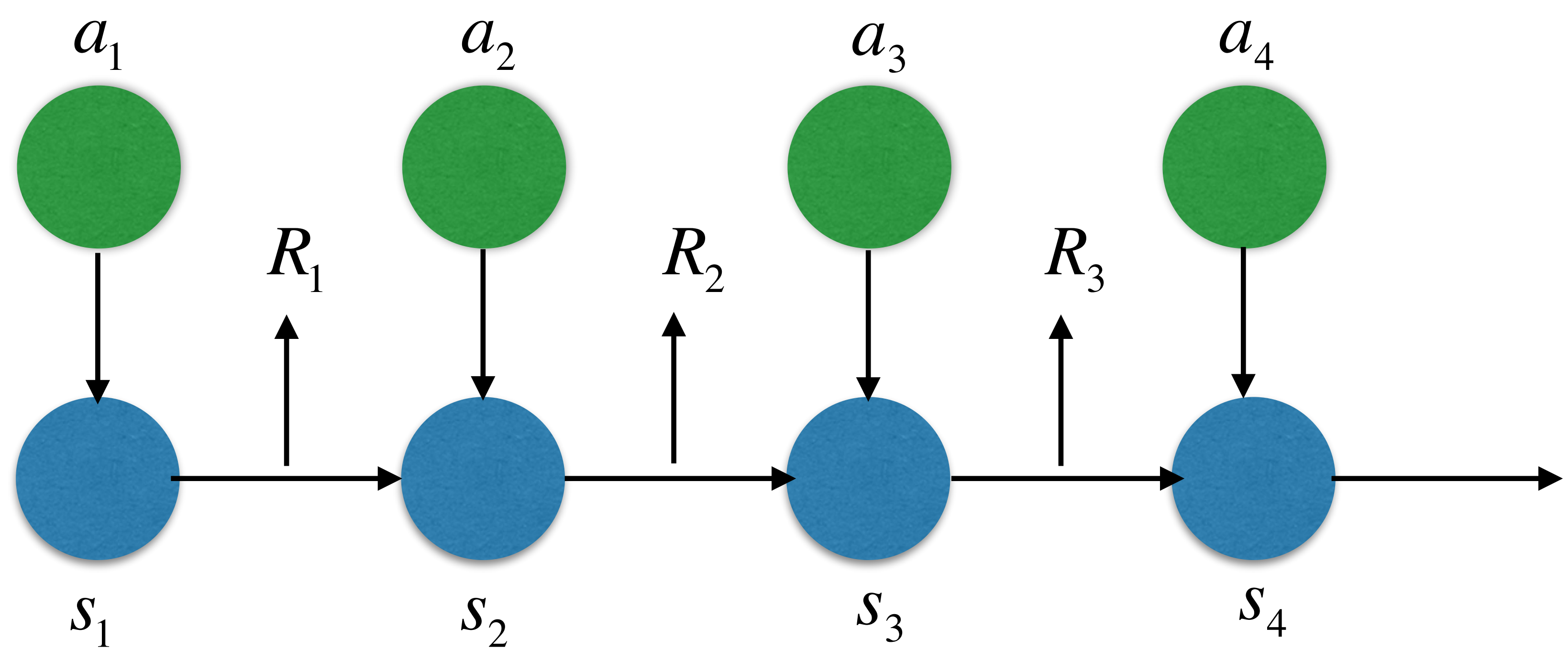
- The **goal** in Reinforcement Learning is to **maximize the expected total reward**

$$\mathbb{E}\left[R(s_0,a_0)+\gamma R(s_1,a_1)+\gamma^2 R(s_2,a_2)+\ldots\right]=\mathbb{E}\left[\sum_n \gamma^n R(s_n,a_n)\right]$$

- This is **risk-neutral** RL (looks only at a mean of the distribution of total reward
- **Risk-sensitive** RL looks at **risk** (e.g. the variance of the total reward) as well…

Decision policy

$$R(s_0,a_0)+\gamma R(s_1,a_1)+\gamma^2 R(s_2,a_2)+\ldots=\sum_n \gamma^n R(s_n,a_n)$$



- The **goal** in Reinforcement Learning is to **maximize the expected total reward**

$$\mathbb{E}\left[R(s_0,a_0)+\gamma R(s_1,a_1)+\gamma^2 R(s_2,a_2)+\ldots\right]=\mathbb{E}\left[\sum_n \gamma^n R(s_n,a_n)\right]$$

- This is achieved by an **optimal choice of a policy** $\pi : S \mapsto A$
- Whenever in state s_t, we take action $a_t = \pi(s_t)$
- Policy π can be deterministic or stochastic (then $\pi(s_t)$ is a probability distribution.

Decision policy

- The goal in Reinforcement Learning is to maximize the expected total reward

$$\mathbb{E}\left[R(s_0,a_0)+\gamma R(s_1,a_1)+\gamma^2 R(s_2,a_2)+\ldots\right]=\mathbb{E}\left[\sum_n \gamma^n R(s_n,a_n)\right]$$

- The value function for policy π

$$V^\pi(s)=\mathbb{E}\left[R(s_0,a_0)+\gamma R(s_1,a_1)+\gamma^2 R(s_2,a_2)+\ldots \mid s_0=s,\pi\right]$$

- The Bellman equation for value function

$$\boxed{V^\pi(s)=R(s)+\gamma\sum_{s'\in S} p(s'\mid s,a=\pi(s))V^\pi(s')}$$

- The Bellman equation is exactly solvable for discrete sets S as a system of $|S|$ linear equations.

Control question

Select all correct answers

1. The goal of (risk-neutral) Reinforcement Learning is to maximize the expected total reward by choosing an optimal policy.
2. The goal of (risk-neutral) Reinforcement Learning is to neutralize risk, i.e. make it equal zero.
3. The goal of risk-sensitive Reinforcement Learning is to incorporate some measures of risk of the distribution of total reward, into the optimal decision process.
4. The goal of risk-sensitive Reinforcement Learning can be achieved by randomly adding totally random actions to optimization, so that the result would be more sensitive to risk of any possible model mis-specification.

Correct answers: 1, 3