

Markov Decision Processes

CS60077: Reinforcement Learning

Abir Das

IIT Kharagpur

Sep 14 and 15, 2020



Agenda

- § Understand definitions and notation to be used in the course.
- § Understand definition and setup of sequential decision problems.



Resources

- § Reinforcement Learning by David Silver [[Link](#)]
- § Deep Reinforcement Learning by Sergey Levine [[Link](#)]
- § SB: Chapter 3

Terminology and Notation

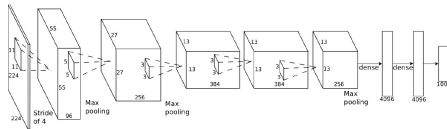
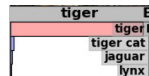
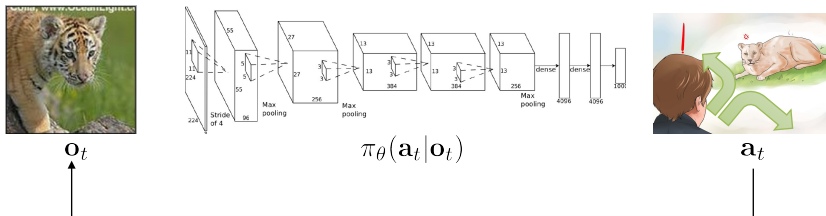
 \mathbf{o}_t  $\pi_{\theta}(\mathbf{a}_t | \mathbf{o}_t)$  \mathbf{a}_t

Figure credit: *S. Levine - CS 294-112 Course, UC Berkeley*

Terminology and Notation



\mathbf{s}_t – state

\mathbf{o}_t – observation

\mathbf{a}_t – action

$\pi_{\theta}(\mathbf{a}_t | \mathbf{o}_t)$ – policy

$\pi_{\theta}(\mathbf{a}_t | \mathbf{s}_t)$ – policy (fully observed)



\mathbf{s}_t – state

\mathbf{o}_t – observation

Figure credit: S. Levine - CS 294-112 Course, UC Berkeley

Student Markov Reward Process

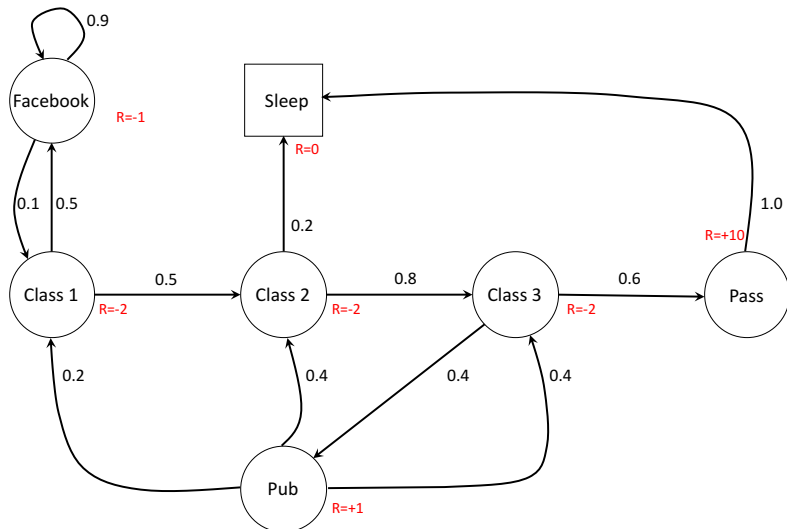


Figure credit: David Silver, DeepMind

Return

Definition

The return G_t is the total discounted reward from timestep t .

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \quad (1)$$

- § $\gamma \in [0, 1]$ is the discounted present value of the future rewards.
- § Immediate rewards are valued above delayed rewards.
 - ▶ γ close to 0 leads to “myopic” evaluation.
 - ▶ γ close to 1 leads to “far-sighted” evaluation.

Why Discount?

Most Markov reward and decision processes are discounted. Why?

§ Uncertainty about the future may not be fully represented

Why Discount?

Most Markov reward and decision processes are discounted. Why?

- § Uncertainty about the future may not be fully represented
- § Immediate rewards are valued above delayed rewards.
- § Avoids infinite returns in cyclic Markov processes or infinite horizon problems.

Value Function

The value function $v(s)$ gives the long-term value of state s

Definition

The *state value function* $v(s)$ of an MRP is the expected return starting from state s

$$v(s) = \mathbb{E}[G_t | S_t = s] \quad (2)$$

Example Student MRP Returns

Sample **returns** for Student MRP:
Starting from $S_1 = C1$ with $\gamma = \frac{1}{2}$

$$G_1 = R_2 + \gamma R_3 + \dots + \gamma^{T-1} R_{T+1}$$

- § C1 C2 C3 Pass Sleep
- § C1 FB FB C1 C2 Sleep
- § C1 C2 C3 Pub C2 C3 Pass Sleep
- § C1 FB FB C1 C2 C3 Pub C1 FB
FB FB C1 C2 C3 Pub C2 Sleep

- § $-2 - \frac{1}{2} * 2 - \frac{1}{4} * 2 + \frac{1}{8} * 10 = -2.25$
- § $-2 - \frac{1}{2} * 1 - \frac{1}{4} * 1 - \frac{1}{8} * 2 - \frac{1}{16} * 2 = -3.125$
- § $-2 - \frac{1}{2} * 2 - \frac{1}{4} * 2 + \frac{1}{8} * 1 - -\frac{1}{16} * 2 - \frac{1}{32} * 2 + \frac{1}{64} * 10 = -3.41$
- § $-2 - \frac{1}{2} * 1 - \frac{1}{4} * 1 - \frac{1}{8} * 2 - \frac{1}{16} * 2 + \dots = -3.20$

State-Value Function for Student MRP (2)

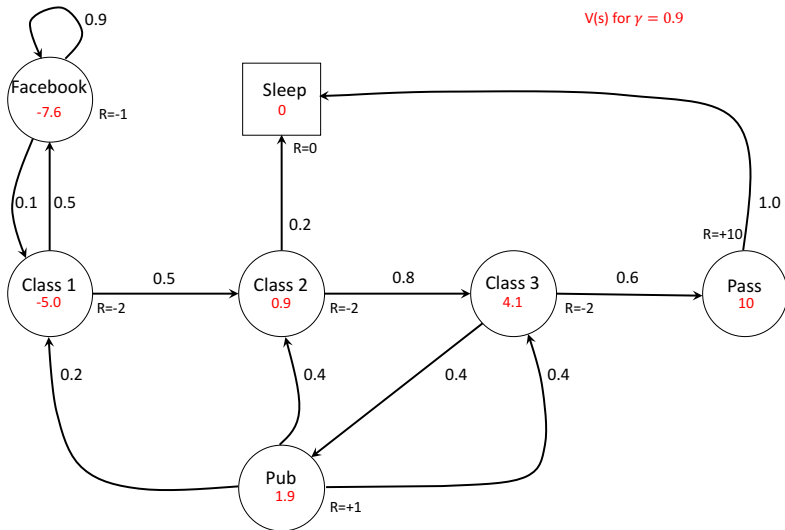


Figure credit: David Silver, DeepMind

State-Value Function for Student MRP (3)

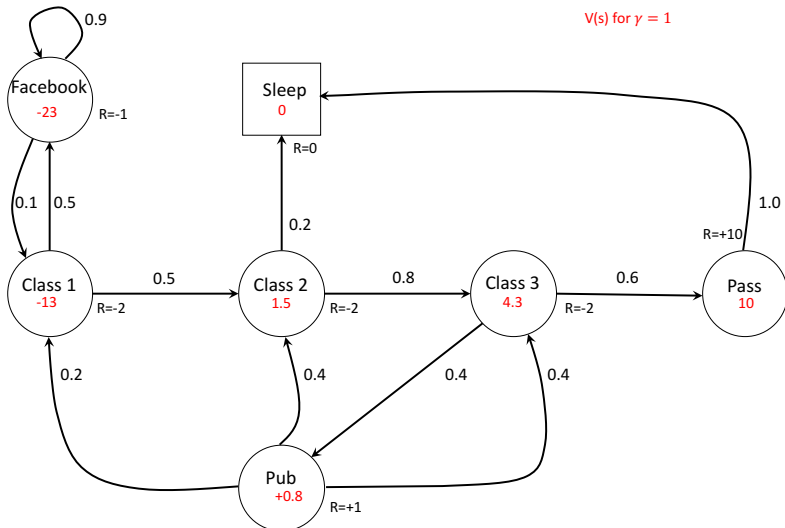


Figure credit: David Silver, DeepMind

Bellman Equation for MRPs

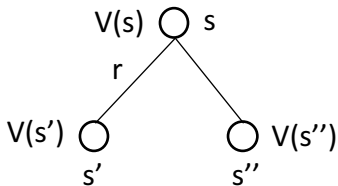
The value function can be decomposed into two parts:

- § immediate reward $R(s)$
- § discounted value of successor state $\gamma v(s')$

$$\begin{aligned}
 v(s) &= R(s) + \gamma \mathbb{E}_{s' \in \mathcal{S}} [v(s')] \\
 &= R(s) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'} v(s') \quad (3)
 \end{aligned}$$



Richard Bellman



Bellman Equation for MRPs - Proof

$$v(s) = \mathbb{E}[G_t | S_t = s] = \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \dots | S_t = s]$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 v(s) &= \mathbb{E}[G_t | S_t = s] = \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \dots | S_t = s] \\
 &= \mathbb{E}[R_{t+1}(S_t) + \gamma R_{t+2}(S_{t+1}) + \gamma^2 R_{t+3}(S_{t+2}) + \gamma^3 R_{t+4}(S_{t+3}) + \dots | S_t = s]
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 v(s) &= \mathbb{E}[G_t | S_t = s] = \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \dots | S_t = s] \\
 &= \mathbb{E}[R_{t+1}(S_t) + \gamma R_{t+2}(S_{t+1}) + \gamma^2 R_{t+3}(S_{t+2}) + \gamma^3 R_{t+4}(S_{t+3}) + \dots | S_t = s] \\
 &= \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+1}(S_t) + \gamma R_{t+2}(S_{t+1}) + \right. \\
 &\qquad \qquad \qquad \left. \gamma^2 R_{t+3}(S_{t+2}) + \gamma^3 R_{t+4}(S_{t+3}) + \dots] \right)
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 v(s) &= \mathbb{E}[G_t | S_t = s] = \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \dots | S_t = s] \\
 &= \mathbb{E}[R_{t+1}(S_t) + \gamma R_{t+2}(S_{t+1}) + \gamma^2 R_{t+3}(S_{t+2}) + \gamma^3 R_{t+4}(S_{t+3}) + \dots | S_t = s] \\
 &= \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+1}(S_t) + \gamma R_{t+2}(S_{t+1}) + \right. \\
 &\qquad\qquad\qquad \left. \gamma^2 R_{t+3}(S_{t+2}) + \gamma^3 R_{t+4}(S_{t+3}) + \dots] \right) \\
 &= \sum_{S_{t+1}, S_{t+2}, \dots} P(S_{t+1}, S_{t+2}, \dots | S_t = s) R_{t+1}(S_t) + \\
 &\qquad \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\qquad\qquad\qquad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right)
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) \sum_{S_{t+1}, S_{t+2}, \dots} P(S_{t+1}, S_{t+2}, \dots | S_t = s) + \\
 &\quad \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right)
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) \sum_{S_{t+1}, S_{t+2}, \dots} P(S_{t+1}, S_{t+2}, \dots | S_t = s) + \\
 &\quad \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right)
 \end{aligned}$$

Bellman Equation for MDPs - Proof

$$\begin{aligned}
&= R_{t+1}(S_t) \sum_{S_{t+1}, S_{t+2}, \dots} P(S_{t+1}, S_{t+2}, \dots | S_t = s) + \\
&\quad \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
&\quad \quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \\
&= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+1}, S_{t+2}, \dots | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
&\quad \quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \\
&= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}, S_t = s) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \right. \\
&\quad \quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right)
\end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \text{ [Conditional independence (Ref eq. (7))]}
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \text{ [Conditional independence (Ref eq. (7))]} \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right)
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots \right] \text{[Conditional independence (Ref eq. (7))]} \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots \right] \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} | S_t = s) \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots \right]
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \text{[Conditional independence (Ref eq. (7))]} \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} | S_t = s) \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} | S_t = s) v(S_{t+1})
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \text{[Conditional independence (Ref eq. (7))]} \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} | S_t = s) \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots] \right) \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} | S_t = s) v(S_{t+1}) \\
 &= R_{t+1}(S_t = s) + \gamma \sum_{s' \in \mathcal{S}} P(S_{t+1} = s' | S_t = s) v(S_{t+1} = s')
 \end{aligned}$$

Bellman Equation for MRPs - Proof

$$\begin{aligned}
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}, S_{t+2}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \gamma R_{t+3}(S_{t+2}) + \right. \\
 &\quad \left. \gamma^2 R_{t+4}(S_{t+3}) + \dots \right] \text{[Conditional independence (Ref eq. (7))]} \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) P(S_{t+1} | S_t = s) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots \right] \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} | S_t = s) \sum_{S_{t+2}, S_{t+3}, \dots} \left(P(S_{t+2}, \dots | S_{t+1}) [R_{t+2}(S_{t+1}) + \right. \\
 &\quad \left. \gamma R_{t+3}(S_{t+2}) + \gamma^2 R_{t+4}(S_{t+3}) + \dots \right] \\
 &= R_{t+1}(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} | S_t = s) v(S_{t+1}) \\
 &= R_{t+1}(S_t = s) + \gamma \sum_{s' \in \mathcal{S}} P(S_{t+1} = s' | S_t = s) v(S_{t+1} = s') = R(s) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'} v(s')
 \end{aligned}$$

Bellman Equation in Matrix Form

So, we have seen,

$$v(s) = R(s) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'} v(s')$$

Where are the time subscripts? **Hint:** Think about (1). Definition of value function, (2). Expectation operation.

Bellman Equation in Matrix Form

So, we have seen,

$$v(s) = R(s) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'} v(s')$$

Where are the time subscripts? **Hint:** Think about (1). Definition of value function, (2). Expectation operation.

The Bellman equation can be expressed concisely using matrices.

$$\mathbf{v} = \mathcal{R} + \gamma \mathcal{P} \mathbf{v}$$

where \mathbf{v} and \mathcal{R} are column vectors with one entry per state.

$$\begin{bmatrix} v(s_1) \\ v(s_2) \\ \vdots \\ v(s_n) \end{bmatrix} = \begin{bmatrix} R(s_1) \\ R(s_2) \\ \vdots \\ R(s_n) \end{bmatrix} + \gamma \begin{bmatrix} \mathcal{P}_{11} & \mathcal{P}_{12} & \cdots & \mathcal{P}_{1n} \\ \mathcal{P}_{21} & \mathcal{P}_{22} & \cdots & \mathcal{P}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{P}_{n1} & \mathcal{P}_{n2} & \cdots & \mathcal{P}_{nn} \end{bmatrix} \begin{bmatrix} v(s_1) \\ v(s_2) \\ \vdots \\ v(s_n) \end{bmatrix}$$

Solving Bellman Equation

- § The Bellman equation being a linear equation, it can be solved directly.

$$\begin{aligned}\mathbf{v} &= \mathcal{R} + \gamma\mathcal{P}\mathbf{v} \\ (\mathbf{I} - \gamma\mathcal{P})\mathbf{v} &= \mathcal{R} \\ \mathbf{v} &= (\mathbf{I} - \gamma\mathcal{P})^{-1}\mathcal{R}\end{aligned}$$

- § As computational complexity is $O(n^3)$ for n states, direct solution is only feasible for small MRPs.
- § There are many iterative methods for large MRPs, e.g., Dynamic programming, Monte-Carlo, Temporal difference learning

Existence of Solution to Bellman Equation

- § We need to show that $(\mathbf{I} - \gamma\mathcal{P})$ is invertible and for that we will use the following result from linear algebra - The inverse of a matrix exists if and only if all its eigenvalues are non-zero.

Existence of Solution to Bellman Equation

- § We need to show that $(\mathbf{I} - \gamma\mathcal{P})$ is invertible and for that we will use the following result from linear algebra - The inverse of a matrix exists if and only if all its eigenvalues are non-zero.
- § For a stochastic matrix (row sum equal to 1 and all entries are ≥ 0), the largest eigenvalue is 1.

Existence of Solution to Bellman Equation

- § We need to show that $(\mathbf{I} - \gamma\mathcal{P})$ is invertible and for that we will use the following result from linear algebra - The inverse of a matrix exists if and only if all its eigenvalues are non-zero.
- § For a stochastic matrix (row sum equal to 1 and all entries are ≥ 0), the largest eigenvalue is 1.

Proof

As \mathcal{P} is a stochastic matrix, $\mathcal{P}\mathbf{1} = \mathbf{1}$ where $\mathbf{1} = [1, 1, \dots, 1]^T$. This means 1 is an eigenvalue of \mathcal{P} .

Now, let's suppose $\exists \lambda > 1$ and non-zero \mathbf{x} such that $\mathcal{P}\mathbf{x} = \lambda\mathbf{x}$.

Since the rows of \mathcal{P} are non-negative and sum to 1, each element of vector $\mathcal{P}\mathbf{x}$ is a convex combination of the components of the vector \mathbf{x} .

A convex combination can't be greater than x_{\max} , the largest component of \mathbf{x} . However, as $\lambda > 1$, at least one element (λx_{\max}) in the R.H.S. (i.e., in $\lambda\mathbf{x}$) is greater than x_{\max} . This is a contradiction and so $\lambda > 1$ is not possible.

Existence of Solution to Bellman Equation

§ So the largest eigenvalue of \mathcal{P} is 1.

Existence of Solution to Bellman Equation

§ So the largest eigenvalue of \mathcal{P} is 1.

Theorem and its proof

For all eigenvalues λ_i of a square matrix \mathbf{A} and corresponding eigenvectors \mathbf{v}_i such that $\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$,

$$\text{eig}(\mathbf{I} + \gamma\mathbf{A}) = 1 + \gamma\lambda_i \quad [\gamma \text{ is any scalar}]$$

Proof:

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$$

$$\gamma\mathbf{A}\mathbf{v}_i = \gamma\lambda_i\mathbf{v}_i$$

$$\mathbf{v}_i + \gamma\mathbf{A}\mathbf{v}_i = \mathbf{v}_i + \gamma\lambda_i\mathbf{v}_i$$

$$(\mathbf{I} + \gamma\mathbf{A})\mathbf{v}_i = (1 + \gamma\lambda_i)\mathbf{v}_i$$

Existence of Solution to Bellman Equation

§ So the largest eigenvalue of \mathcal{P} is 1.

Theorem and its proof

For all eigenvalues λ_i of a square matrix \mathbf{A} and corresponding eigenvectors \mathbf{v}_i such that $\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$,

$$\text{eig}(\mathbf{I} + \gamma\mathbf{A}) = 1 + \gamma\lambda_i \quad [\gamma \text{ is any scalar}]$$

Proof:

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$$

$$\gamma\mathbf{A}\mathbf{v}_i = \gamma\lambda_i\mathbf{v}_i$$

$$\mathbf{v}_i + \gamma\mathbf{A}\mathbf{v}_i = \mathbf{v}_i + \gamma\lambda_i\mathbf{v}_i$$

$$(\mathbf{I} + \gamma\mathbf{A})\mathbf{v}_i = (1 + \gamma\lambda_i)\mathbf{v}_i$$

§ So the smallest eigenvalue of $(\mathbf{I} - \gamma\mathcal{P})$ is $1 - \gamma$. For $\gamma < 1$ which is > 0 . And hence, $(\mathbf{I} - \gamma\mathcal{P})$ is invertible.

Markov Decision Process

A Markov decision process is a Markov reward process with actions.

Definition

A Markov Decision Process is a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, where

§ \mathcal{S} is the state space (can be continuous or discrete)

§ \mathcal{A} is the action space (can be continuous or discrete)

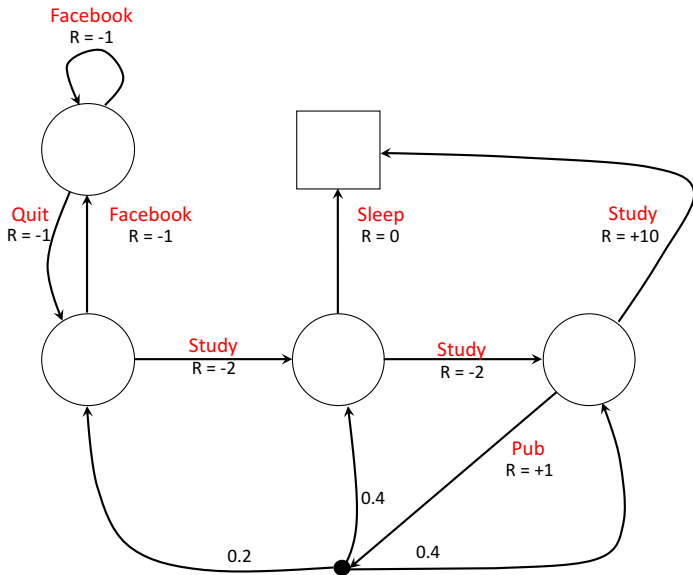
§ \mathcal{P} is the state transition probability matrix.

$$\mathcal{P}_{ss'}^a = P(S_{t+1} = s' | S_t = s, A_t = a) = p(s'/s, a)$$

§ \mathcal{R} is a reward function, $\mathcal{R} = \mathbb{E}[R_{t+1} | S_t = s, A_t = a] = R(s, a)$

§ γ is a discount factor, $\gamma \in [0, 1]$

Example: Student MDP



Policy

Definition

A *policy* π is a distribution over actions given states,

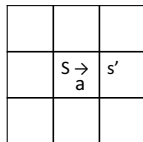
$$\pi(a/s) = P[A_t = a | S_t = s]$$

- § The Markov property means the policy depends on the current state (not the history)
- § The policy can be either deterministic or stochastic
- § The policy can be either stationary or non-stationary

Policy

§ For a deterministic environment $p(s'/s, a) = 1$,
else for stochastic environment $0 \leq p(s'/s, a) \leq 1$

§ In a stochastic environment, there is always some chance to end up in s' starting from state s and taking any action.



§ So, probability of ending up in state s' from s irrespective of the action (*i.e.*, taking any action according to the policy), = probability of taking action 1 from state $s \times$ probability of ending up in state s' taking action 1 + probability of taking action 2 from state $s \times$ probability of ending up in state s' taking action 2 + \dots

§ This means $p_\pi(s'|s) = \sum_a \pi(a|s)p(s'|s, a)$

§ Similarly, the one-step expected reward for following policy π is given by $r_\pi(s) = \sum_a \pi(a|s)r(s, a)$

§ Side note: The above is given by $r_\pi(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|s, a)r(s, a, s')$ when reward is a function of the transiting state s' also.

Value Functions

Definition

The *state-value* function $v_\pi(s)$ of an MDP is the expected return starting from state s , and then following policy π

$$v_\pi(s) = \mathbb{E}_\pi [G_t | S_t = s] \quad (4)$$

Definition

The *action-value* function $q_\pi(s, a)$ of an MDP is the expected return starting from state s , taking action a , and then following policy π

$$q_\pi(s, a) = \mathbb{E}_\pi [G_t | S_t = s, A_t = a] \quad (5)$$

Example: State-Value function for Student MDP

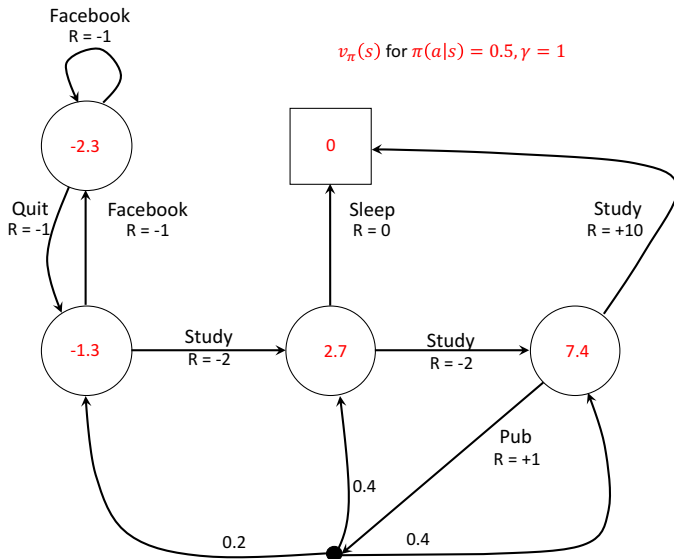
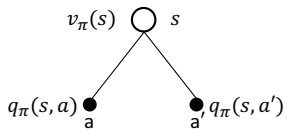


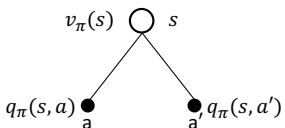
Figure credit: David Silver, DeepMind

Relation between v_π and q_π

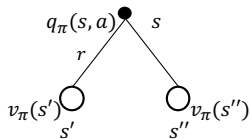


$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) q_\pi(s, a)$$

Relation between v_π and q_π

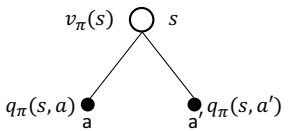


$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) q_\pi(s, a)$$

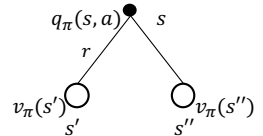


$$q_\pi(s, a) = r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_\pi(s')$$

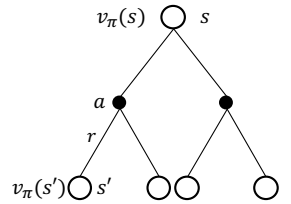
Relation between v_π and q_π



$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s)q_\pi(s, a)$$

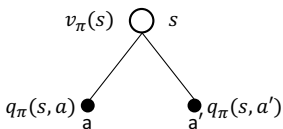


$$q_\pi(s, a) = r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a)v_\pi(s')$$

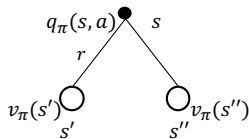


$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left\{ r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a)v_\pi(s') \right\}$$

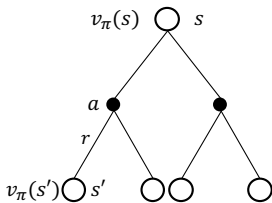
Relation between v_π and q_π



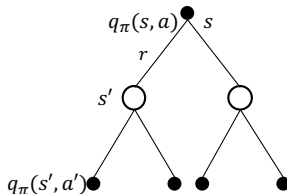
$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) q_\pi(s, a)$$



$$q_\pi(s, a) = r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_\pi(s')$$



$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left\{ r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_\pi(s') \right\}$$



$$q_\pi(s, a) = r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) \left\{ \sum_{a' \in \mathcal{A}} \pi(a'|s') q_\pi(s', a') \right\}$$

Bellman Expectation Equations

Like MRPs, the value function can be decomposed into two parts - immediate reward $r(s)$ and the discounted value of successor state $\gamma v(s')$. But, as action is involved in MDP, the form is a little different.

$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \sum_{s' \in \mathcal{S}} p(s'|s, a) \{r(s, a, s') + \gamma v_\pi(s')\}$$

[when r is a function of s, a, s']

$$= \sum_{a \in \mathcal{A}} \pi(a|s) \left\{ r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_\pi(s') \right\}$$

[when r is a function of s, a]

$$= r(s) + \gamma \sum_{a \in \mathcal{A}} \pi(a|s) \sum_{s' \in \mathcal{S}} p(s'|s, a) v_\pi(s')$$

[when r is a function of s]

(6)

Bellman Expectation Equations

$$\begin{aligned}q_\pi(s, a) &= \mathbb{E}_\pi[G_t | S_t = s, a_t = a] \quad [\text{eqn. 3.13 in SB}] \\&= \mathbb{E}_\pi[r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots | S_t = s, a_t = a] \\&= \mathbb{E}_\pi[r_{t+1} + \gamma(r_{t+2} + \gamma r_{t+3} \dots) | S_t = s, a_t = a] \\&= \mathbb{E}_\pi[r_{t+1} + \gamma G_{t+1} | S_t = s, a_t = a] \quad [\text{By definition, eqn. 3.11 in SB}] \\&= \mathbb{E}_\pi[r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi[G_{t+1} | S_t = s, a_t = a]\end{aligned}$$

Bellman Expectation Equations

$$\begin{aligned}
 q_{\pi}(s, a) &= \mathbb{E}_{\pi} [G_t | S_t = s, a_t = a] \quad [\text{eqn. 3.13 in SB}] \\
 &= \mathbb{E}_{\pi} [r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots | S_t = s, a_t = a] \\
 &= \mathbb{E}_{\pi} [r_{t+1} + \gamma(r_{t+2} + \gamma r_{t+3} \dots) | S_t = s, a_t = a] \\
 &= \mathbb{E}_{\pi} [r_{t+1} + \gamma G_{t+1} | S_t = s, a_t = a] \quad [\text{By definition, eqn. 3.11 in SB}] \\
 &= \mathbb{E}_{\pi} [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_{\pi} [G_{t+1} | S_t = s, a_t = a] \\
 &= \mathbb{E}_{\pi} [r_{t+1} | S_t = s, a_t = a] + \\
 &\quad \gamma \mathbb{E}_{\pi} [\mathbb{E}_{\pi} [G_{t+1} | S_t = s, a_t = a, S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a]
 \end{aligned}$$

(Above applies the formula $\mathbb{E}[Y|X] = \mathbb{E}[\mathbb{E}[Y|X, Z]|X]$)

[Get the intuition behind the formula in [this youtube link](#)]

Bellman Expectation Equations

$$\begin{aligned}
 q_\pi(s, a) &= \mathbb{E}_\pi [G_t | S_t = s, a_t = a] \quad [\text{eqn. 3.13 in SB}] \\
 &= \mathbb{E}_\pi [r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots | S_t = s, a_t = a] \\
 &= \mathbb{E}_\pi [r_{t+1} + \gamma(r_{t+2} + \gamma r_{t+3} \dots) | S_t = s, a_t = a] \\
 &= \mathbb{E}_\pi [r_{t+1} + \gamma G_{t+1} | S_t = s, a_t = a] \quad [\text{By definition, eqn. 3.11 in SB}] \\
 &= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a] \\
 &= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \\
 &\quad \gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a, S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]
 \end{aligned}$$

(Above applies the formula $\mathbb{E}[Y|X] = \mathbb{E}[\mathbb{E}[Y|X, Z]|X]$)

[Get the intuition behind the formula in [this](#) youtube link]

$$\begin{aligned}
 &= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \\
 &\quad \gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]
 \end{aligned}$$

[G_{t+1} depends only on s_{t+1} and a_{t+1}]

Bellman Expectation Equations

$$q_\pi(s, a) = \mathbb{E}_\pi [G_t | S_t = s, a_t = a] \quad [\text{eqn. 3.13 in SB}]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma(r_{t+2} + \gamma r_{t+3} \dots) | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma G_{t+1} | S_t = s, a_t = a] \quad [\text{By definition, eqn. 3.11 in SB}]$$

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] +$$

$$\gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a, S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]$$

(Above applies the formula $\mathbb{E}[Y|X] = \mathbb{E}[\mathbb{E}[Y|X, Z]|X]$)

[Get the intuition behind the formula in [this](#) youtube link]

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] +$$

$$\gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]$$

[G_{t+1} depends only on s_{t+1} and a_{t+1}]

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi [q_\pi(s', a') | S_t = s, a_t = a] \quad [\text{Using definition of } q_\pi]$$

Bellman Expectation Equations

$$q_\pi(s, a) = \mathbb{E}_\pi [G_t | S_t = s, a_t = a] \quad [\text{eqn. 3.13 in SB}]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma(r_{t+2} + \gamma r_{t+3} \dots) | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma G_{t+1} | S_t = s, a_t = a] \quad [\text{By definition, eqn. 3.11 in SB}]$$

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] +$$

$$\gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a, S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]$$

(Above applies the formula $\mathbb{E}[Y|X] = \mathbb{E}[\mathbb{E}[Y|X, Z]|X]$)

[Get the intuition behind the formula in [this](#) youtube link]

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] +$$

$$\gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]$$

[G_{t+1} depends only on s_{t+1} and a_{t+1}]

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi [q_\pi(s', a') | S_t = s, a_t = a] \quad [\text{Using definition of } q_\pi]$$

Bellman Expectation Equations

$$q_\pi(s, a) = \mathbb{E}_\pi [G_t | S_t = s, a_t = a] \quad [\text{eqn. 3.13 in SB}]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma(r_{t+2} + \gamma r_{t+3} \dots) | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} + \gamma G_{t+1} | S_t = s, a_t = a] \quad [\text{By definition, eqn. 3.11 in SB}]$$

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a]$$

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] +$$

$$\gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_t = s, a_t = a, S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]$$

(Above applies the formula $\mathbb{E}[Y|X] = \mathbb{E}[\mathbb{E}[Y|X, Z]|X]$)

[Get the intuition behind the formula in [this](#) youtube link]

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] +$$

$$\gamma \mathbb{E}_\pi \left[\mathbb{E}_\pi [G_{t+1} | S_{t+1} = s', a_{t+1} = a'] | S_t = s, a_t = a \right]$$

[G_{t+1} depends only on s_{t+1} and a_{t+1}]

$$= \mathbb{E}_\pi [r_{t+1} | S_t = s, a_t = a] + \gamma \mathbb{E}_\pi [q_\pi(s', a') | S_t = s, a_t = a] \quad [\text{Using definition of } q_\pi]$$

Bellman Expectation Equations

$$= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a', s' | s, a)$$

$$= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s', s, a) p(s' | s, a)$$

$$= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s') p(s' | s, a) \text{ [Markov property]}$$

Bellman Expectation Equations

$$= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a', s' | s, a)$$

$$= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s', s, a) p(s' | s, a)$$

$$= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s') p(s' | s, a) \text{ [Markov property]}$$

$$= r(s, a) + \sum_{s' \in \mathcal{S}} p(s' | s, a) \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s')$$

Bellman Expectation Equations

$$\begin{aligned}
 &= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a', s' | s, a) \\
 &= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s', s, a) p(s' | s, a) \\
 &= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s') p(s' | s, a) \text{ [Markov property]} \\
 &= r(s, a) + \sum_{s' \in \mathcal{S}} p(s' | s, a) \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s')
 \end{aligned}$$

Bellman Expectation Equations

$$\begin{aligned}
 &= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a', s' | s, a) \\
 &= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s', s, a) p(s' | s, a) \\
 &= r(s, a) + \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s') p(s' | s, a) \text{ [Markov property]} \\
 &= r(s, a) + \sum_{s' \in \mathcal{S}} p(s' | s, a) \sum_{a' \in \mathcal{A}} q_{\pi}(s', a') p(a' | s')
 \end{aligned}$$

Bellman Expectation Equation for Student MDP

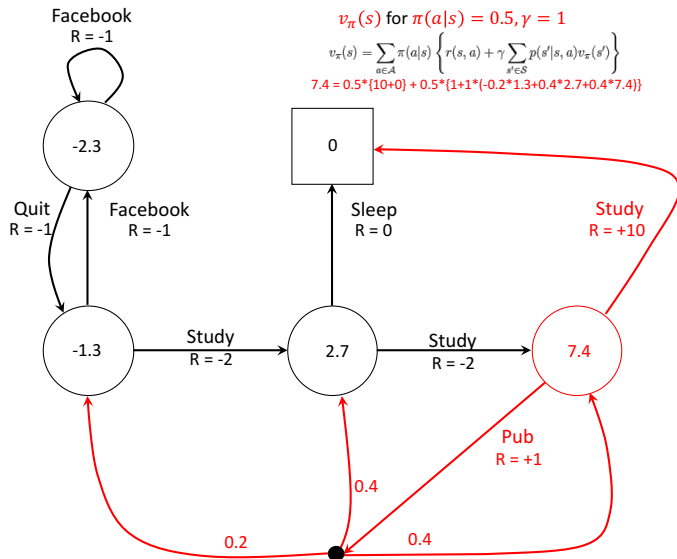


Figure credit: David Silver, DeepMind

Optimal Policies and Optimal Value Functions

- § Solving a reinforcement learning task means, roughly, finding a policy that achieves a lot of reward (*maximum*) over the long run.
- § The notion of maximality leads to *optimality* in MDPs.

Optimal Policies and Optimal Value Functions

- § Solving a reinforcement learning task means, roughly, finding a policy that achieves a lot of reward (*maximum*) over the long run.
- § The notion of maximality leads to *optimality* in MDPs.
- § What is meant by a policy is better than some other policy?

Optimal Policies and Optimal Value Functions

- § Solving a reinforcement learning task means, roughly, finding a policy that achieves a lot of reward (*maximum*) over the long run.
- § The notion of maximality leads to *optimality* in MDPs.
- § What is meant by a policy is better than some other policy?
- § A policy π is defined to be better than or equal to a policy π' if its expected return is greater than or equal to that of π' for all states.

Definition

$$\pi \geq \pi' \text{ iff } v_\pi(s) \geq v_{\pi'}(s), \forall s \in \mathcal{S}$$

Optimal Policies and Optimal Value Functions

Definition

The *optimal* state-value function $v_*(s)$ is the maximum state-value function over all policies

$$v_*(s) = \max_{\pi} v_{\pi}(s), \forall s \in \mathcal{S}$$

The *optimal* action-value function $q_*(s, a)$ is the maximum action-value function over all policies

$$q_*(s, a) = \max_{\pi} q_{\pi}(s, a), \forall s \in \mathcal{S} \text{ and } a \in \mathcal{A}$$

§ An MDP is “solved” when we know the optimal value function

Optimal Action-Value Function for Student MDP

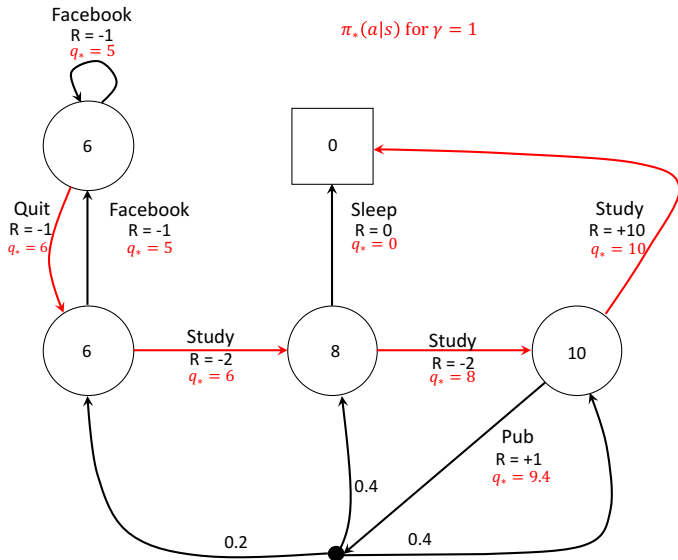


Figure credit: David Silver, DeepMind

Optimal Policy

Theorem

For any Markov Decision Process

- § There exists an optimal policy π_* that is better than or equal to all other policies, $\pi_* \geq \pi, \forall \pi$
- § All optimal policies achieve the optimal value function $v_{\pi_*}(s) = v_*(s)$
- § All optimal policies achieve the optimal action-value function $q_{\pi_*}(s, a) = q_*(s, a)$

Optimal Policy

Theorem

For any Markov Decision Process

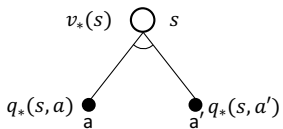
- § There exists an optimal policy π_* that is better than or equal to all other policies, $\pi_* \geq \pi, \forall \pi$
- § All optimal policies achieve the optimal value function $v_{\pi_*}(s) = v_*(s)$
- § All optimal policies achieve the optimal action-value function $q_{\pi_*}(s, a) = q_*(s, a)$

An optimal policy can be found by maximising over $q_*(s, a)$.

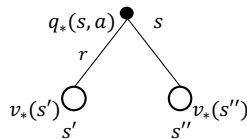
$$\pi_*(a|s) = \begin{cases} 1 & \text{if } a = \arg \max_{a \in \mathcal{A}} q_*(s, a) \\ 0 & \text{otherwise} \end{cases}$$

- § There is always a deterministic optimal policy for any MDP.
- § If we know $q_*(s, a)$, we immediately have the optimal policy.

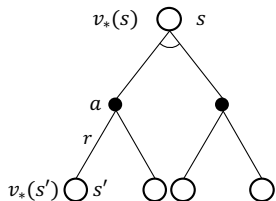
Relation between v_* and q_*



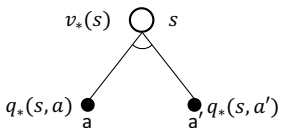
$$v_*(s) = \max_{a \in \mathcal{A}} q_*(s, a)$$



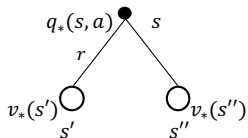
$$q_*(s, a) = r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_*(s')$$



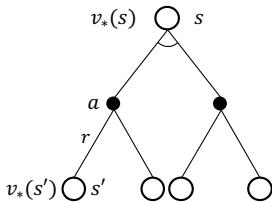
$$v_*(s) = \max_{a \in \mathcal{A}} \left\{ r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_*(s') \right\}$$

Relation between v_* and q_* 

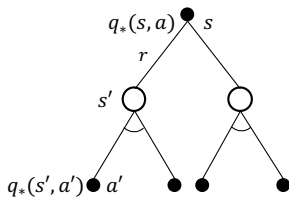
$$v_*(s) = \max_{a \in \mathcal{A}} q_*(s, a)$$



$$q_*(s, a) = r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_*(s')$$



$$v_*(s) = \max_{a \in \mathcal{A}} \left\{ r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) v_*(s') \right\}$$



$$q_*(s, a) = r(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s, a) \max_{a' \in \mathcal{A}} q_*(s', a')$$

Appendices

Appendices

1. Independence

Independence

$$A \perp\!\!\!\perp B \implies P(A|B) = P(A)$$

Conditional Independence

$$A \perp\!\!\!\perp B|C \implies P(A|B, C) = P(A|C)$$

Proof:

$$\begin{aligned}
 P(A|B, C) &= \frac{P(A, B, C)}{P(B, C)} = \frac{P(A, B|C)P(C)}{P(B|C)P(C)} & (7) \\
 &= \frac{P(A|C)P(B|C)}{P(B|C)} \quad [\text{From definition of conditional independence}] \\
 &= P(A|C)
 \end{aligned}$$

2. Independence

Theorem

Eigenvalues of the transpose A^T are the same as the eigenvalues of A

Proof

Eigenvalues of a matrix are roots of its characteristic polynomial. Hence if the matrices A and A^T have the same characteristic polynomial, then they have the same eigenvalues.

$$\begin{aligned}\det(A^T - \lambda I) &= \det(A^T - \lambda I^T) & (8) \\ &= \det(A - \lambda I)^T \\ &= \det(A - \lambda I) \quad [\text{Since } \det(A) = \det(A^T)]\end{aligned}$$