高级算法
Advanced Topics in Algorithms

陈 旭

数据科学与计算机学院

中山大學
SUN YAT-SEN UNIVERSITY

Chapter 2:
Simple Decision Processes

Outline

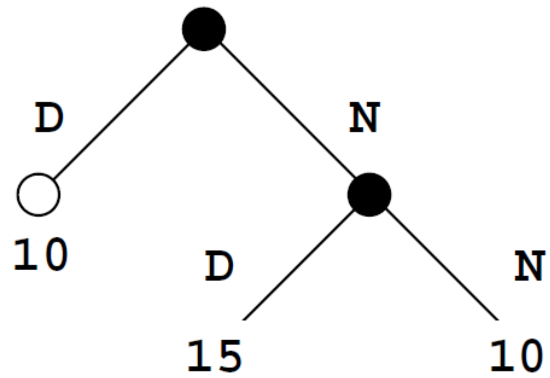
Decision Tree

Motivating Example

- Let say each day, I may ask you to make a decision: I will offer you \$1 or \$10. Which will you take?
- Strategic Behavior: some interesting observations
  - immediate reward are forgone in the expectation of a payback in the future.
  - behavior of others are taken into account.

Decision Tree

(a) The times at which decisions are made are shown as small, filled circle. (b) Leading away from these decision nodes is a branch for every action. (c) Whenever *every* decisions have been made, one reaches the end of one path. Payoff for following the path is written.



Optimal decision

Take nickel (N) first, then take dime (D).

Definition

A *strategy* is a rule for choosing an action at every point that a decision might have to be made. A *pure strategy* is one in which there is no randomization. The set of all possible pure strategies is denoted as $\boldsymbol{S}$.

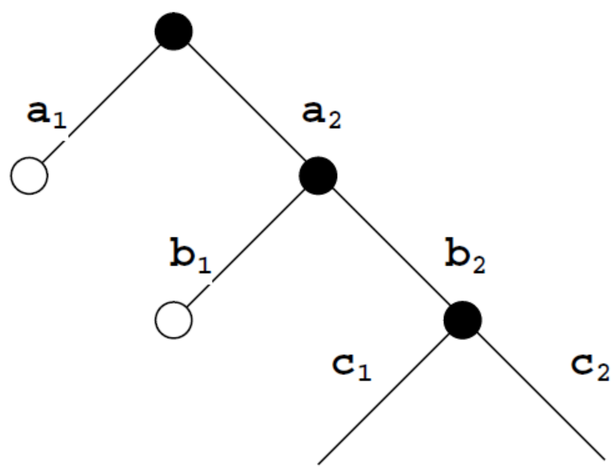
Suppose there are n decision nodes and $\boldsymbol{A}_i$ denote the action set at node i. Some or all of the sets $\boldsymbol{A}_i$ may be identical. The set of pure strategies $\boldsymbol{S} = \boldsymbol{A}_1 \times \boldsymbol{A}_2 \times \cdots \times \boldsymbol{A}_n$.

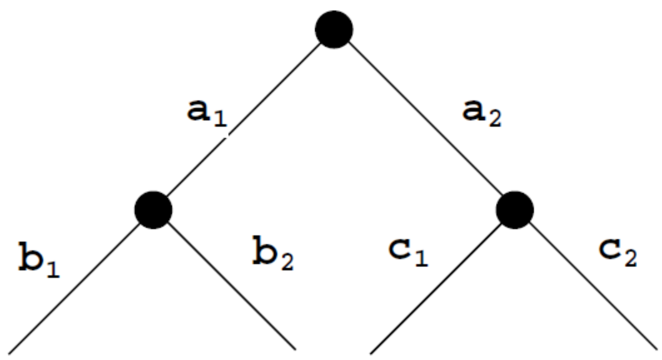
Example

Suppose there are three decision nodes with which $\boldsymbol{A}_1 = \{a_1, a_2\}, \boldsymbol{A}_2 = \{b_1, b_2\}, \boldsymbol{A}_3 = \{c_1, c_2\}$. We have:

$$\boldsymbol{S} = \{a_1b_1c_1, a_1b_1c_2, a_1b_2c_1, a_1b_2c_2, a_2b_1c_1, a_2b_1c_2, a_2b_2c_1, a_2b_2c_2\}.$$

In this example, $\boldsymbol{S}$ could be apply to either of the decision trees.





Definition

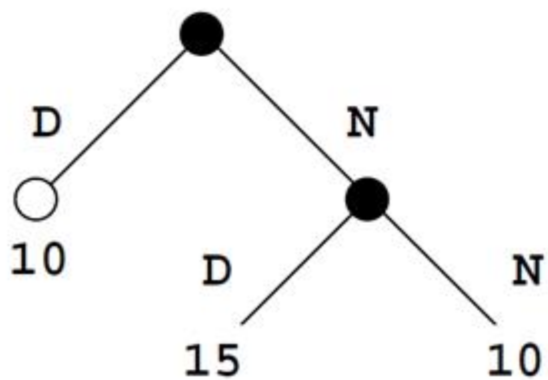
The observed behavior of an individual following a given strategy is called the *outcome* of the strategy.

Notes

- The definition of a strategy leads to some redundancy in terms of outcomes.
- On the one hand, a pure strategy can be viewed as a path from the initial node to a terminal node in the decision tree.
- On the other hand, a pure strategy specifies the action that would be taken at *every* decision nodes, including those that will not be reached if the strategy is followed.
- So observe behavior (outcome) only provides us with a *part* of the strategy.

Example

- Consider the first figure of "nickel or dime" example.
- We have $\boldsymbol{S} = \{DD, DN, ND, NN\}$.
- Note that DD and DN have the same outcome: getting \$10 since the game terminates after choosing D.



Example

- Consider the first figure of “nickel or dime” example.
- We have $\boldsymbol{S} = \{DD, DN, ND, NN\}$.
- Note that *DD* and *DN* have the same outcome: getting $10 since the game terminates after choosing *D*.

We can have the following useful concept.

Definition

A reduced strategy set is the set formed when all pure strategies that lead to indistinguishable outcomes are combined.

For the same example, the reduced strategy set $\boldsymbol{S}_R = \{NN, ND, DX\}$ where *DX* means choosing dime first and anything at the other decision nodes.

When there is only a single decision to be made, the *set of actions* and *pure strategies* are the *same*. Suppose the action (or pure strategy) set is $\{a_1, a_2\}$. The only way of specifying randomizing behavior is to use a_1 with probability p and a_2 with probability $1 - p$. We denote $\beta = (p, 1 - p)$.

Definition

A mixed strategy σ specifies the probability $p(s)$ with which each of the pure strategies $s \in \boldsymbol{S}$.

Suppose the set $\boldsymbol{S} = \{s_a, s_b, s_c, \ldots\}$, then a mixed strategy can be represented as

$$\sigma = (p(s_a), p(s_b), p(s_c), \ldots).$$

A pure strategy can also be represented as a probability vector:

$$s_b = (0, 1, 0, \ldots).$$

Mixed strategies, can then be represented as a *linear combination* of pure strategies:

$$\sigma = \sum_{s \in \boldsymbol{S}} p(s)s.$$

In the "nickel or dime" game, the mixed strategy of playing *NN* with probability $1/4$ and *DN* with probability $3/4$ is:

$$\sigma = \frac{1}{4}NN + \frac{3}{4}DN.$$

Definition

The *support* of a mixed strategy σ is that set $\boldsymbol{S}(\sigma) \subseteq \boldsymbol{S}$ of all the pure strategies for with σ specifies $p(s) > 0$.

Definition

Let the decision nodes be labelled by an indicator set $\boldsymbol{I} = \{1, \ldots, n\}$. At each node i, the action set is $\boldsymbol{A}_i = \{a_1^i, a_2^i, \ldots, a_{k_i}^i\}$. An individual's behavior at node i is determined by the probability vector $\boldsymbol{p}_i = (p(a_1^i), p(a_2^i), \ldots, p(a_{k_i}^i))$. A behavioral strategy β is the collection of probability vectors:

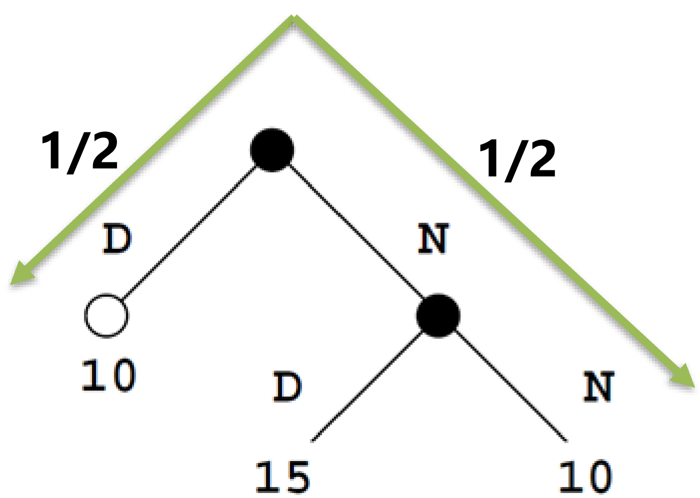
$$\beta = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n\}.$$

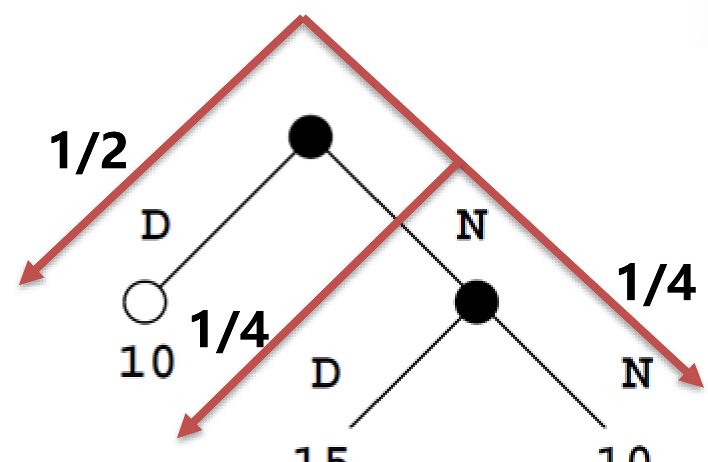
Difference between σ and β

Consider the "nickel or dime" game in the first figure. One mixed strategy is $\sigma = \frac{1}{2}NN + \frac{1}{2}DD$. Is it correct to say that the behavioral strategy is $\beta = ((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))$?

Difference between σ and β

Consider the "nickel or dime" game in the first figure. One mixed strategy is $\sigma = \frac{1}{2}NN + \frac{1}{2}DD$. Is it correct to say that the behavioral strategy is $\beta = ((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))$? **No**. To see why. Note that there are three paths through the decision tree, which we call "dime only", "all nickels" and "nickel and dime". The mixed strategy picks the paths "dime only" and "all nickels" each with probability 1/2 and "nickel and dime" with probability zero. The behavioral strategy would picked the path "nickel and then dime" with probability 1/4 and not zero. So $\sigma \neq \beta$.





Definition

A behavioral strategy and a mixed strategy are *equivalent* if they assign the same probabilities to each of the possible pure strategies that are available. When they are equivalent, they have the same payoff.

Equivalence of σ and β

In the "nickel or dime" game. If $\sigma = \frac{1}{2}NN + \frac{1}{2}DD$, then

- The equivalent $\beta = ((\frac{1}{2}, \frac{1}{2}), (0, 1))$.
- Furthermore, any of the mixed strategies:

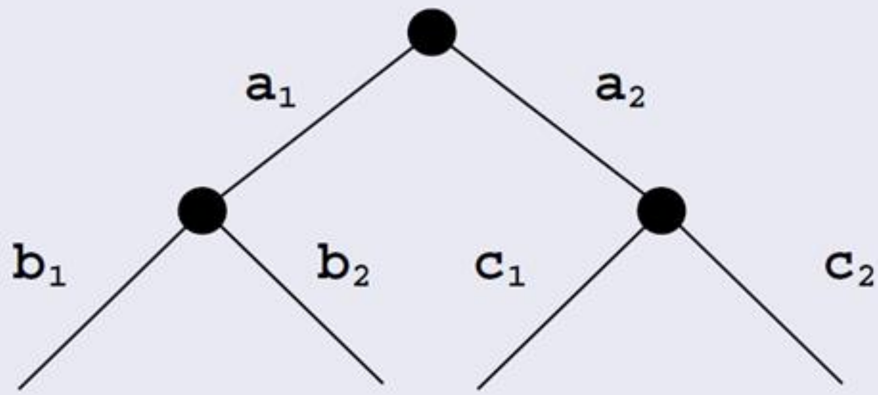
$$\sigma_x = \frac{1}{2}NN + \left(\frac{1}{2} - x\right) DD + xDN \quad \text{with } x \in [0, 1/2]$$

is equivalent to the behavioral strategies $\beta = ((\frac{1}{2}, \frac{1}{2})(0, 1))$.

Theorem

(a) Every behavioral strategy has a mixed representation and (b) every mixed strategy has a behavioral representation.

Find all behavioral strategy equivalents for the mixed strategies (a) $\sigma = \frac{1}{2}a_1b_1c_1 + \frac{1}{2}a_2b_2c_2$ and (b) $\sigma = \frac{1}{3}a_1b_1c_1 + \frac{1}{3}a_1b_2c_1 + \frac{1}{3}a_1b_1c_2$.



(a) $\beta = \left(\left(\frac{1}{2}, \frac{1}{2}\right), (1,0), (0,1)\right)$

(b) $\beta = \left((1,0), \left(\frac{2}{3}, \frac{1}{3}\right), (x, 1-x)\right)$ with $x \in [0,1]$

In previous lecture, we saw the randomizing behavior was not required for single decisions, in the sense that an optimal action could always be found. Similar results hold for decision processes.

Theorem

Let σ^* be an optimal mixed strategy with support $\boldsymbol{S}^*$. Then $\pi(s) = \pi(\sigma^*)$ $\forall s \in \boldsymbol{S}^*$.

Proof

If $|\boldsymbol{S}^*| = 1$, then it is obviously true. Let say $|\boldsymbol{S}^*| \geq 2$. If theorem is false, then at least one $s' \in \boldsymbol{S}^*$ gives the highest payoff than $\pi(\sigma^*)$ (we prove by contradiction), then

$$\begin{aligned}\pi(\sigma^*) &= \sum_{s \in \boldsymbol{S}^*} p^*(s)\pi(s) = \sum_{s \neq s'} p^*(s)\pi(s) + p^*(s')\pi(s') \\ &< \sum_{s \neq s'} p^*(s)\pi(s') + p^*(s')\pi(s') = \pi(s')\end{aligned}$$

which contradicts that the original assumption that σ^* is optimal.

Theorem

For any decision process, an optimal pure strategy can always be found.

Proof

Previously, we show that every behavioral strategy has at least one equivalent mixed strategy. It follows that no behavioral strategy can have a payoff greater than the corresponding mixed strategy. Therefore, based on the previous theorem, if an optimal strategy exists, then an optimal pure strategy also exists.

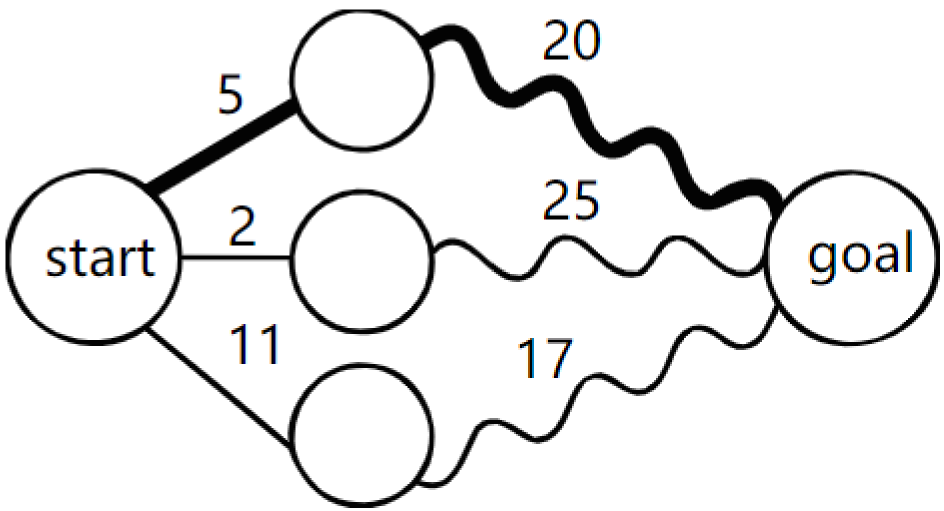
Insight

This implies a procedure to find optimal strategy: list the possible pure strategies, evaluate their payoffs, pick the optimal. But this can be computational expensive. If a tree has n nodes and each node has two actions, there are 2^n pure strategies.

Principle of Optimality

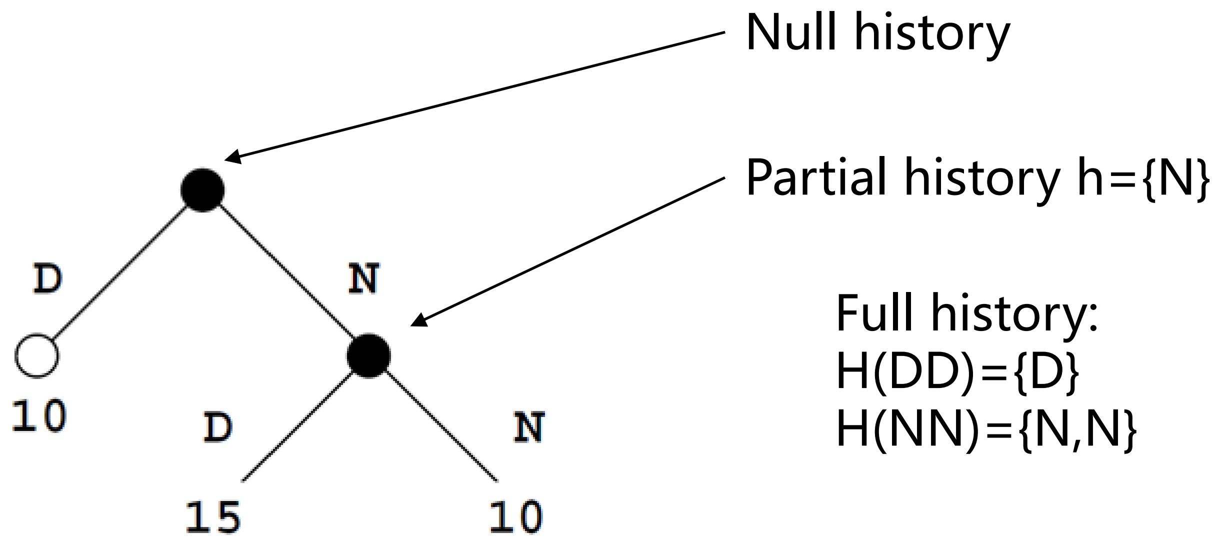
Intuitive idea

To reduce complexity, rely on the Principle of Optimality: at any point along the optimal path, the remaining path is optimal. Therefore, to find the optimal decision *now*, we should assume that we will behave optimally in the *future*.



Definition

A partial history h is the sequence of decision that have been made by an individual up to some specific time. At the start of a decision process (when no decision has been made), we have the null history, $h = \emptyset$. A full history for a strategy s is the complete sequence of all decisions that would be made by an individual following s and is denoted as $H(s)$.



Full history:
H(DD)={D}
H(NN)={N,N}

Definition

Define the subset of pure strategies $S(h) \in S$ that contains all the strategies with history h but that differ in that actions taken in the future. Then the optimal payoff an individual can achieve given that the history h is

$$\pi^*(s|h) = \max_{s \in S(h)} \pi(s).$$

comment

Assume that the individual now has a choice from a set of action $A(h)$. After that decision has been made, the history will be the sequence h with the chosen action a appended, denoted as h, a.

Theorem

For an individual with perfect recall (e.g., he remembers all the past decisions), then:

1. $\pi^*(s|H(s)) = \pi(s)$
2. $\pi^*(s|h) = \max_{a \in A(h)} \pi^*(s|h, a)$
3. $\pi^* = \max_{a \in S(\emptyset)} \pi^*(s|\emptyset)$.

Proof

1. By the definition of $H(s)$, the individual has no more decision to make and the best payoff they can get is the payoff they have already achieved by using strategy s.

Proof: continue

2. A pure strategy is a sequence of actions $\{a_0, a_1, \ldots, a_h, a_{h+1}, \ldots, a_H\}$. So

$$\pi(s) = \pi(a_0, a_1, \ldots, a_h, a_{h+1}, \ldots, a_H).$$

   Let the partial history h be the given sequence $\{a_0, a_1, \ldots, a_h\}$, then

$$\begin{aligned}\pi^*(s|h) &= \max_{a_{h+1}} \max_{a_{h+2}} \ldots \max_{a_H} \pi(a_0, a_1, \ldots, a_h, a_{h+1}, \ldots, a_H) \\ &= \max_{a_{h+1}} \pi^*(s|h, a_{h+1}).\end{aligned}$$

3. The history $h = \emptyset$ denote the optimization problem starting from the beginning, so $S(\emptyset) = S$ and

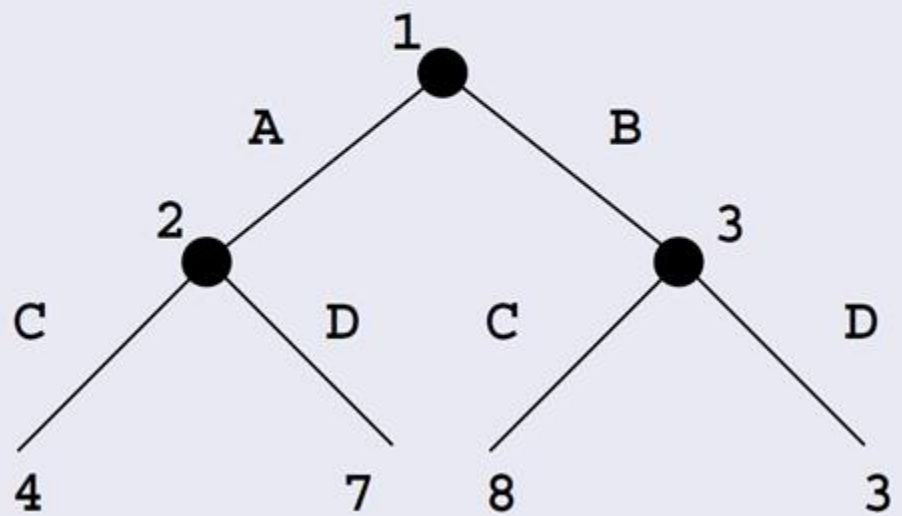
$$\max_{s \in S(\emptyset)} \pi^*(s|\emptyset) = \max_{s \in S} \pi(s) = \pi^*.$$

Key idea:

We should work *backwards* through the decision tree, or what we called the backward induction.

Example

- Determine the optimal strategy for the following decision tree.

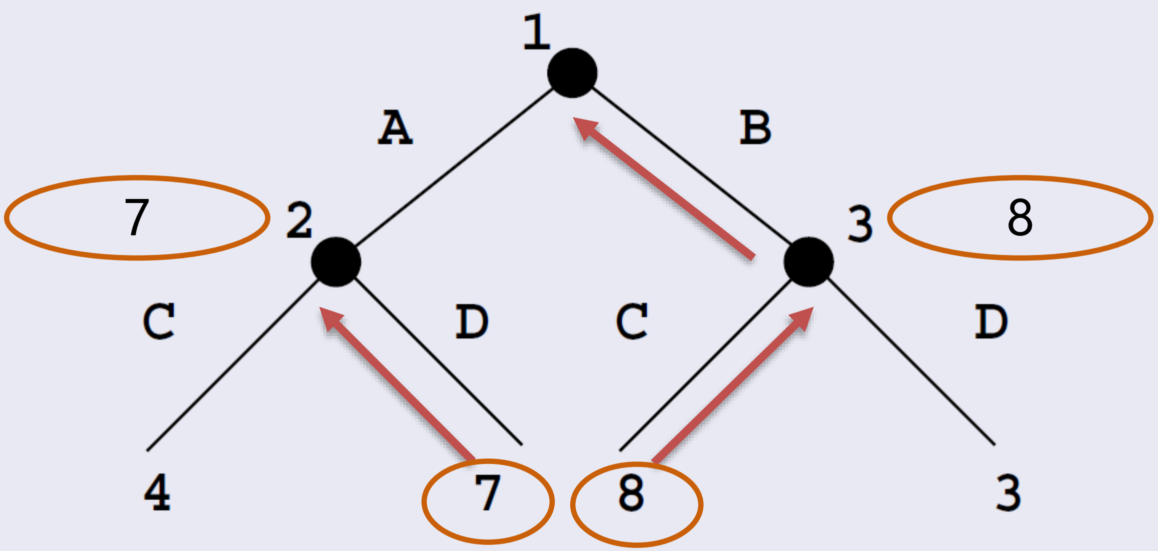


Key idea:

We should work *backwards* through the decision tree, or what we called the backward induction.

Example

- Determine the optimal strategy for the following decision tree.



- **Answer:** BDC (in the order of the labelling of the decision nodes).

Assignments

- P18, Exercise 1.7
- P34, Exercise 2.1
- P42, Exercise 2.4

- 提交邮箱: gjsf_2020@126.com
- 邮件+文件命名：学号+姓名
- 提交期限：下周四23:59