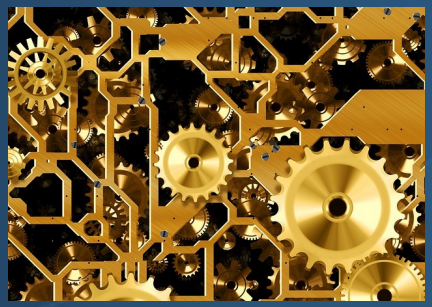
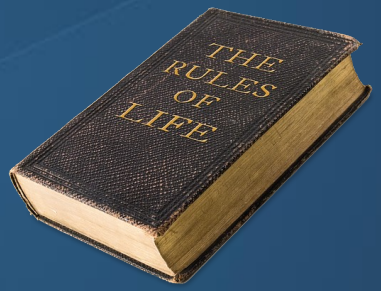
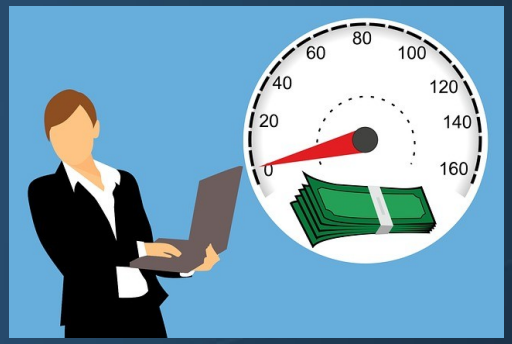
Deep Q Learning: From Paper to Code

Model Free vs. Model Based Learning

Last Time







Markov property + policy → Value function

$$v_\pi(s) = \sum_a \pi(a,s) \sum_{s',r} p(s',r|s,a)[r + \gamma v_\pi(s')]$$

Goodness of state

Bellman Equation

A Problem?

$$v_\pi(s) = \sum_a \pi(a,s) \sum_{s',r} p(s',r|s,a)[r + \gamma v_\pi(s')]$$

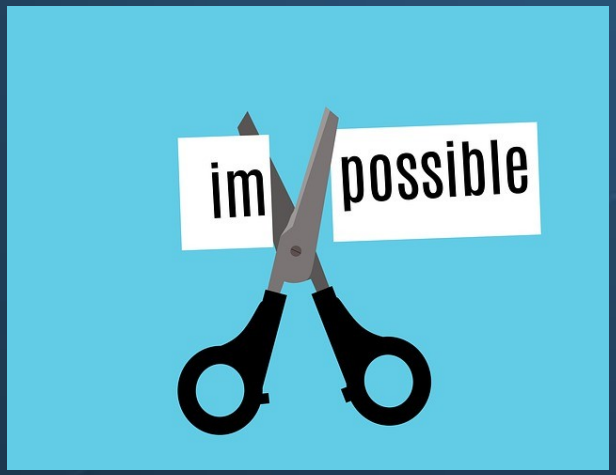
How to handle unknown?

No way to solve without p!

Known Probability Distribution

Suppose distribution is known ...

$\Pi(s, a) \rightarrow$ probability of selecting a in s



Success!

Work backwards from terminal state

$G_T = 0$

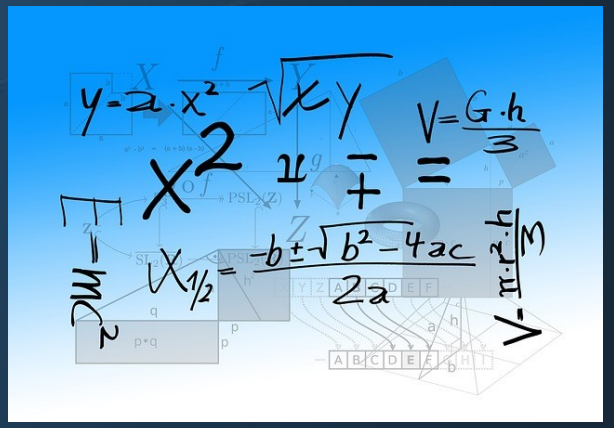
$v_\pi(s_{terminal}) = E_\pi[G_t | S_t = s] = E_\pi[0] = 0$

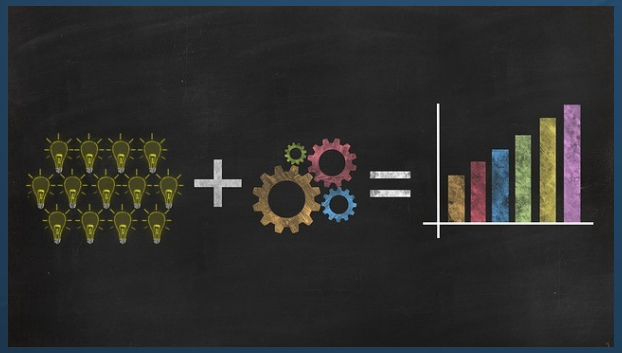
Use recursion relationship

Known Probability Distribution

Complete model of environment

Solve (large) system of equations





Allows explicit planning

Dynamic Programming

Unknown Probability Distribution

Model Free → Q Learning

Estimate p by playing

Much more flexible

Biological Parallels

Use model based and model free

We can learn from maps

We can also learn by trial and error

Summary

- Solve B.E. explicitly or through exploration
- Q learning is model free
- Dynamic programming → solving equations
- Estimating p is critical to our success

Up Next