

# RL基础

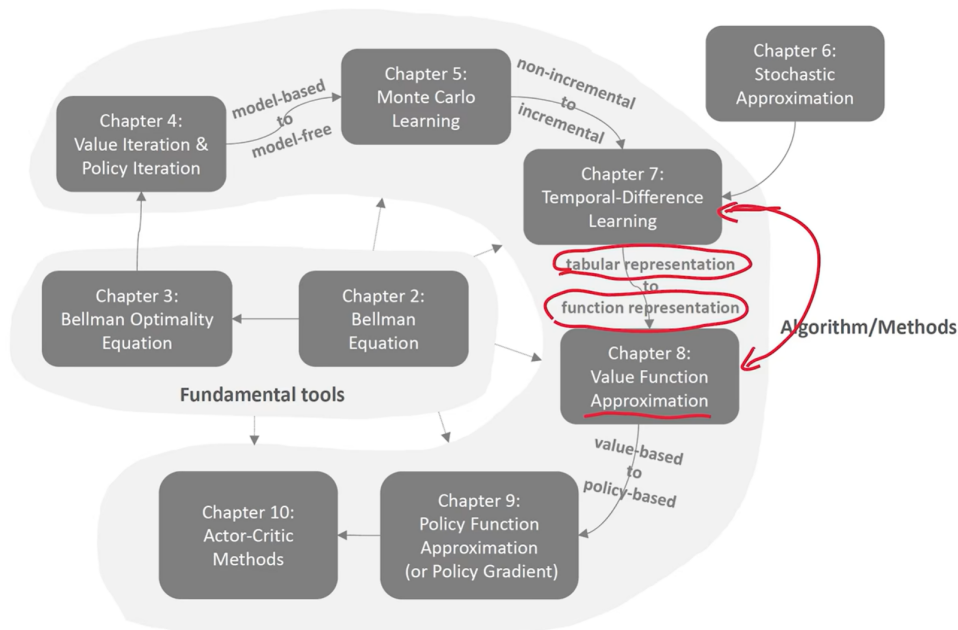
【强化学习的数学原理】课程：从零开始到透彻理解（完结）\_哔哩哔哩\_bilibili

## 8-值函数近似

 Note

第一次引入神经网络;

mindmap



主要内容

- 1 Motivating examples: curve fitting ←
- 2 Algorithm for state value estimation
  - Objective function ←
  - Optimization algorithms ←
  - Selection of function approximators ←
  - Illustrative examples
  - Summary of the story
  - Theoretical analysis
- 3 Sarsa with function approximation
- 4 Q-learning with function approximation
- 5 Deep Q-learning
- 6 Summary

## 引言

- 表格型方法：表格是指action-value的二维的表格表示；
- 缺点：state space或者action space较大、甚至连续时，表格型难以存储、难以泛化；

- 例子

- Suppose there are one-dimensional states  $s_1, \dots, s_{|S|}$ .
- Their state values are  $v_\pi(s_1), \dots, v_\pi(s_{|S|})$ , where  $\pi$  is a given policy.
- Suppose  $|S|$  is very large and we hope to use a simple curve to approximate these dots to save storage.

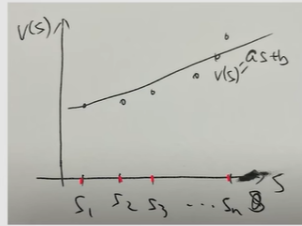


Figure: An illustration of function approximation of samples.

- 一维state数量非常多，使用一个v函数来近似表示各个state的value ( $\hat{v}(s, w) \approx v_\pi(s)$ )；如此，不用存储各个state对应的精确的value，而只需要存储曲线的参数（如线性拟合）；
  - 节省大量的存储；
  - 增强了泛化性；访问一个状态，会改变函数参数，其他state-value也会变化；
- 对于非线性的曲线而言，本质上对于参数w而言仍然是线性变化，对于变量s而言需要先设计一个kernel function进行变换；

$$\hat{v}(s, w) = as^2 + bs + c = \underbrace{[s^2, s, 1]}_{\phi^T(s)} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_w = \phi^T(s)w.$$

- 也可以用神经网络做非线性的拟合；

## state-value 的近似