

Policy vs Plan

Policy vs Plan

$$V(s) = \max_a \left(R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$

From now on, we call this bellman equation

Policy vs Plan












V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
V=0.66	V=0.73	V=0.81	V=0.73

Artificial Intelligence A-Z

© SuperDataScience

This is a deterministic approach.

Policy vs Plan

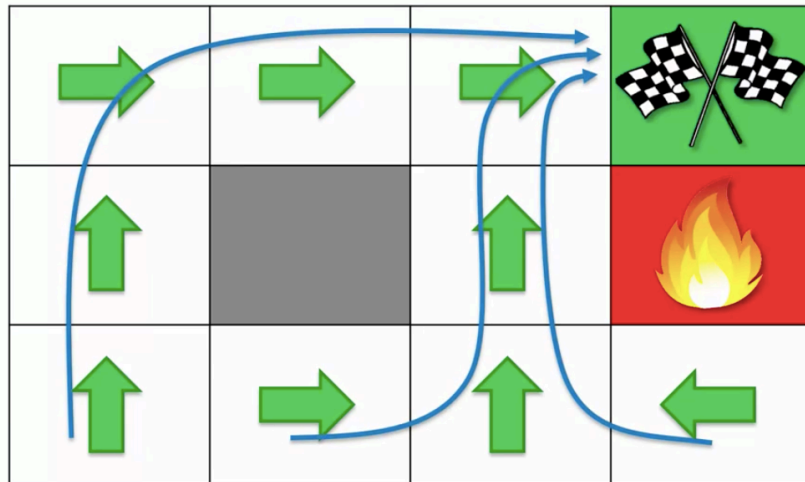
			
			
			

Artificial Intelligence A-Z

© SuperDataScience

This is plan. Sometimes we don't or can't have a plan. For instance, sometimes the plan doesn't happen. Not all the time, things are under plan. For example, when you click to the right but it goes to down.

Policy vs Plan



Artificial Intelligence A-Z

© SuperDataScience

Policy vs Plan

V=0.71	V=0.74	V=0.86	
V=0.63		V=0.39	
V=0.55	V=0.46	V=0.36	V=0.22

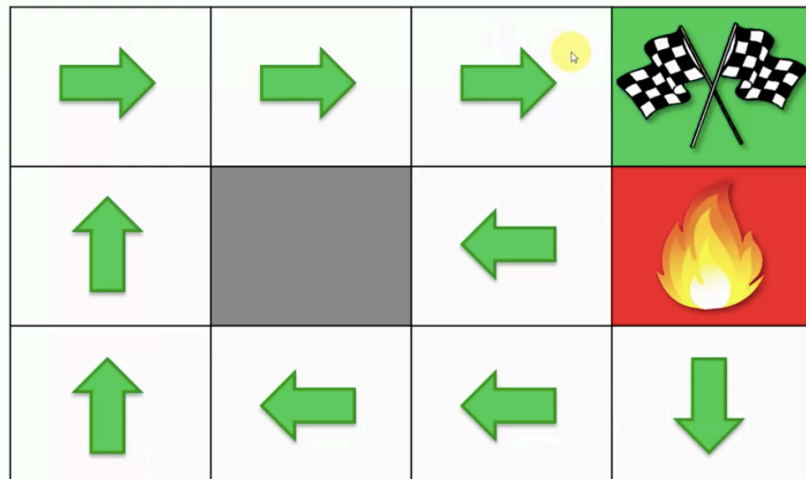
Artificial Intelligence A-Z

© SuperDataScience

This is our new values given that we have randomness inside. For example, $v=0.86$ was equal to 1 previously and the reason for changing it, is because there is 100% deterministic that it goes to the right. For example, there is a chance of 10% that it goes to top to the wall or to the bottom. So to sum up, previously we had a deterministic approach and in here we have non-deterministic values. For finding these values calculate the bellman equation. The reason $V=0.39$ dropped dramatically comparing to other numbers is because it's next to the -1-reward state.

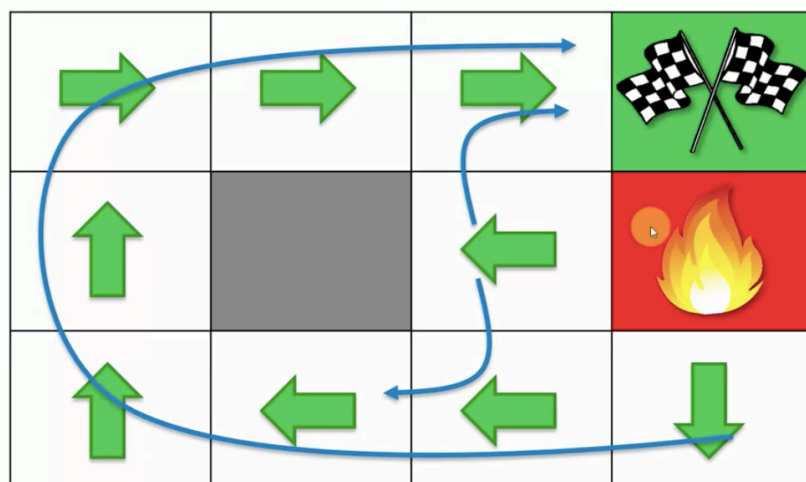
If the agent is at $V=0.22$ the AI encourages it to go from the left path rather than right one because there is a low chance of losing there. (look at the number)

Policy vs Plan



The reason the arrows around the pit are at the opposite direction is because so it first go to the wall (let's say 80%) and for 10% go to right and 10% to the left. With this approach there is a 0% chance to go to the pitfall. This is learnt by experimentation like the self-learning dog.

Policy vs Plan



Policy vs Plan

