



- y-axis: Cost
- x-axis: steps
- curve: average performance

1. What kind of RL algorithms did you use? value-based, policy-based, model-based? why?

value-based

To get the flag on the mountain, the car need the previous observation to decide the next action it should do, and I choose deep Q network, which is a value-based method.

2. This algorithms is off-policy or on-policy? why?

off-policy

deep Q network is belong to off-policy method, it will keep change the action with the reward it got, this method will not get a wrong direction because of an accidental mistake .

3. How does your algorithm solve the correlation problem in the same MDP?

By changing the reward to meet the correlation problem's requirements.