MCTS
Graded Quiz • 8 min

**Due** Nov 18, 8:59 AM EET

✓ **Congratulations! You passed!**

**TO PASS** 80% or higher

Keep Learning

Retake the assignment in **7h 58m**

GRADE
92%

# MCTS

LATEST SUBMISSION GRADE

92%

1. **What is true about planning in RL?** 1 / 1 point

☑ Planning allows to *compute* (contrast with *learn*)the best possible action.

✓ Correct

□ For planning, we do not need to explore – we already know all we need to compute optimal policy.

☑ Planning is computationaly intensive.

✓ Correct

□ Planning does not make use of Dynamic Programming

☑ Planning algorithms that can output a valid policy (or value function) at each moment of planning (even before the end of planning) are called *anytime planning algorithms.*

✓ Correct

2. **What are the differences between model-free and model-based settings?** 1 / 1 point

□ In a model-free setting, we know which (reward, next state) pairs are possible given current state and action.

□ In a model-based setting, we know nothing about environment dynamics. An agent is learning by optimizing some parametric model.

□ Model-free learning requires probabilities of possible next states and rewards but does not require a sample model (aka simulator, aka generative model)

☑ In a model-free setting, we know nothing about environment dynamics. Optimization of agent decisions is based solely on sample-based experiences of the world.

✓ Correct

☑ In a model-based setting, we can find out which (reward, next state) pairs are possible given current state and action.

✓ Correct

3. **What are different types of planning?** 1 / 1 point

☑ Background planning is synonymous to model-free learning from precollected samples of the given environment model with the goal to improve policy / value function in all the sampled states .

✓ Correct

□ Background planning starts after an agent's transition into a new state; it is used to select an optimal action for the current state only.

☑ Decision time planning starts after an agent's transition into a new state; it is used to select an optimal action for the current state only.

✓ **Correct**

☐ Decision time planning is synonymous to model-free learning from precollected samples of the given environment model with the goal to improve policy / value function in all the sampled states .

4. **What is true about *rollout policy* (RP) and *tree policy* (TP) in MCTS?**

**0.8 / 1 point**

☐ MCTS is relatively insensitive to the quality of RP. So we can use random RP, but it is always better to make RP as good as possible.

☑ For long episodes and very short planning time RP should be as fast as possible, TP can be much slower.

✓ **Correct**

☐ Random RP is best, since improving the quality of RP may reduce the quality of MCTS planning.

☐ RP *and* TP are conflicting policies -- the one that performs the best becomes the final result of planning.

☐ For long episodes and short planning time TP should be as fast as possible, RP can be much slower.

You didn't select all the correct answers

5. **What is true about MCTS?**

**0.8 / 1 point**

☐ The backup phase of MCTS performs Policy Evaluation: it makes value estimates in tree nodes consistent with the compound policy (rollout + tree policy)

☑ The MAX action selection strategy (classic MCTS) is an instance of Policy Improvement -- greedy action selection with respect to action-value function.

✓ **Correct**

☐ MCTS uses binary search to select the best action in each node in the tree since it is the fastest strategy out there.

☑ MCTS balances between exploration and exploitation by treating action selection in each node as an independent multiarmed bandit problem.

✓ **Correct**

☐ Preserving action-value function estimates in tree nodes after a transition into a next state is a heuristic: in the new state policy changes, so values are not valid anymore.

You didn't select all the correct answers