Congratulations! You passed!

Grade received 90% | Latest Submission Grade 90% | To pass 80% or higher

Go to next item

1. Which of the following are the most accurate characterizations of sample models and distribution models? (Select all that apply) **0 / 1 point**

- [ ] A sample model can be used to compute the probability of all possible trajectories in an episodic task based on the current state and action.
- [ ] A sample model can be used to obtain a possible next state and reward given the current state and action, whereas a distribution model can only be used to compute the probability of this next state and reward given the current state and action.
- [ ] Both sample models and distribution models can be used to obtain a possible next state and reward, given the current state and action.
- [x] A distribution model can be used as a sample model.

Correct
Correct; a distribution model contains all the information about the transition dynamics of the system, which can be used to 'sample' new states and rewards given the current state and action – just like a sample model.

You didn't select all the correct answers

2. Which of the following statements are TRUE for Dyna architecture? (Select all that apply) **1 / 1 point**

- [x] Real experience can be used to improve the value function and policy

Correct
Correct; we do this in the direct-RL step of the tabular Dyna-Q algorithm

- [x] Simulated experience can be used to improve the value function and policy

Correct
Correct; we do this in the planning step of the tabular Dyna-Q algorithm

- [x] Real experience can be used to improve the model

Correct
Correct; we do this in the model-learning step of the tabular Dyna-Q algorithm

- [ ] Simulated experience can be used to improve the model

3. Mark all the statements that are TRUE for the tabular Dyna-Q algorithm. (Select all that apply) **1 / 1 point**

- [x] The environment is assumed to be deterministic.

Correct
Correct; the algorithm assumes that the environment deterministically transitions to a single next state and reward for a given state-action pair. If the environment is stochastic, the update-model step in its current form would simply overwrite a state-action pair with a different next state and reward transition.