1. Which of the following meta-parameters can be tuned to improve performance of the agent? Performance refers to the cumulative reward the agent would receive *in expectation* across different runs. (Select all that apply) — 1 point

- [x] The step size in the update rule of the learning algorithm (e.g., alpha in Q-learning)
- [ ] Random seed (for the random number generator)
- [x] Number of hidden-layer units in a neural network approximating the value function
- [x] Exploration parameter (e.g., epsilon in e-greedy or the temperature tau in the softmax policy)

2. Suppose a problem that you have formulated as an MDP has k continuous input dimensions. You are considering using tile coding as a function approximator. With T tilings and t tiles per dimension in each tiling, which of the following represent the resultant number of features? (Assume each tiling covers all k dimensions.) — 1 point

- (●) T·t^k
- ( ) T · t / k
- ( ) T · t · k
- ( ) k·T^t

3. Which of the following statements regarding feature-construction methods are TRUE? (Select all that apply) — 1 point

- [x] The feature representation obtained using neural networks changes with time.
- [x] In low-dimensional problems, tile coding is computationally efficient and provides good generalization and discrimination.
- [ ] The feature representation obtained using tile coding changes with time.
- [x] A simple implementation of tile coding leads to memory requirements that might be exponential in the

- [x] A simple implementation of tile coding leads to memory requirements that might be exponential in the number of features.

4. True or False: Adding more hidden layers (of a fixed finite width) increases the representation capacity of neural network. For example, if you have a single-hidden layer neural network with 16 units and nonlinear activations, then adding another layer of 16 units to get a neural network with two hidden layers can represent more functions. 1 point

   - (•) True
   - ( ) False

5. True or False: Adding more hidden layers to a neural network increases the number of parameters needed to be learned. 1 point

   - (•) True
   - ( ) False

6. Which of the following statements regarding the exploration approach are TRUE? (Select all that apply) 1 point

   - [x] A softmax policy is a limited strategy for exploration because it can only be used with action preferences and policy-gradient methods.
   - [ ] Both optimistic initial values and epsilon-greedy exploration can be easily used with neural networks, because they are simple exploration strategies.
   - [x] Optimistic initial values are difficult to maintain when using neural networks as a function approximator.
   - [ ] Epsilon-greedy exploration is difficult to combine with neural networks.

7. Which of the following are TRUE about the softmax temperature parameter tau? 1 point

- [x] For very large tau, the agent's policy is nearly a uniformly-random policy.
- [x] If tau is large, the agent's policy is more stochastic.
- [x] For very small tau, the agent mostly selects the greedy action.
- [ ] Tau does not affect the exploration at all.

8. Which of the following statements are true about activation functions? (Select all that apply) 1 point

- [x] The gradient of flat regions in the range of an activation function w.r.t. the input is zero.
- [x] For inputs of large magnitude, the derivative of the sigmoid and tanh functions are close to zero.
- [ ] Rectified Linear Units (ReLUs) are linear activation functions.
- [ ] Linear activation functions (such as f(x)=x) have derivatives close to zero for inputs of large magnitude.

9. Consider you are using a neural network to approximate the action-value function of a reinforcement learning agent. You decide to use a neural network with two hidden layers. Now you want to choose the activation function for the hidden layers and the output layer. One option is to use a neural network with tanh activation functions in both hidden layers, and a linear activation in the output layer. Another option is to use a neural network with linear activations in both the hidden layers and the output layer. 1 point

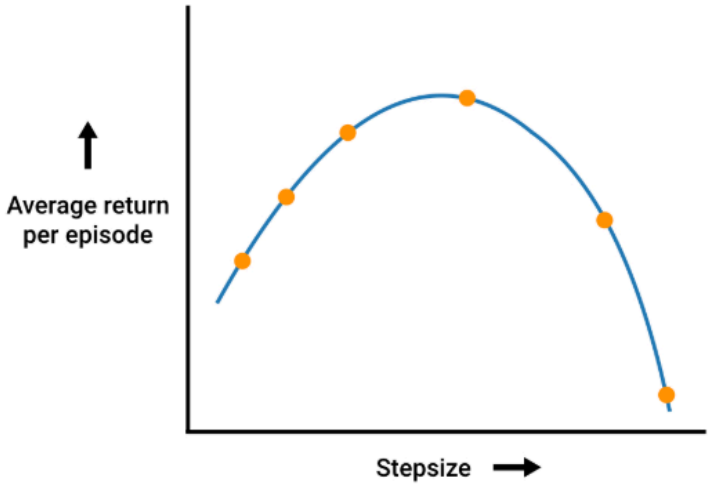
   True or False: In both cases (option one and option two), the neural network can represent the same class of action-value functions.

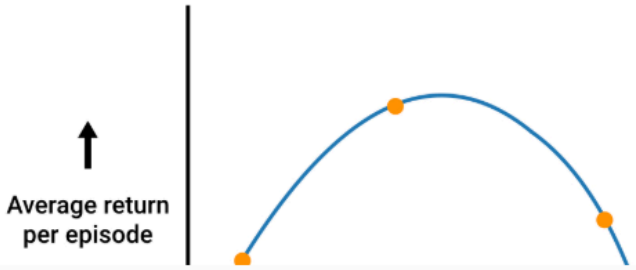
- ( ) True
- (•) False

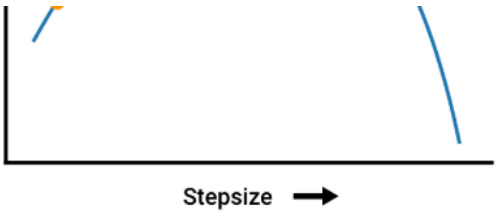
10. Which of the following statements are TRUE regarding methods for selecting a stepsize for the learning update? 1 point

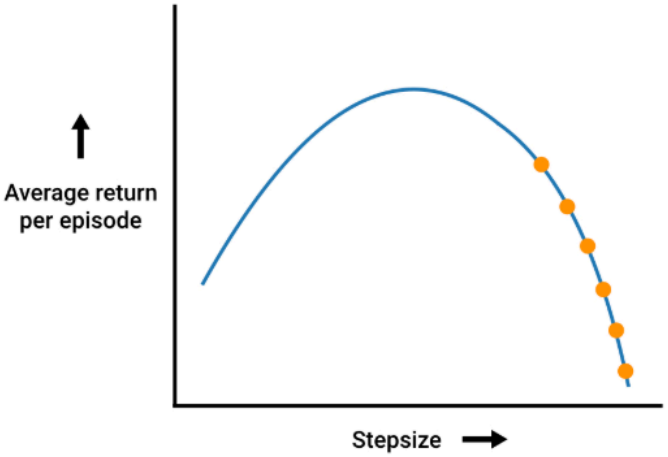
- [x] The heuristic to change the stepsize can be learned from the data collected from the agent-environment interactions.
- [ ] A stepsize that reduces over time (such as 1/N, where N is the number of agent-environment interactions) is necessary when the environment changes over time.
- [x] An adaptive stepsize selection method like RMSProp uses a heuristic to change the stepsize during learning.

11. Suppose we want to find the optimal policy that obtains the maximum undiscounted return per episode in some task. We are using Expected Sarsa. With the rest of the meta-parameters fixed, we want to find the best setting of the stepsize that results in the best performance in this setting. In the following graph, the blue line represents how the performance measure varies with stepsize. Obviously, we do not have this information beforehand, and we are selecting a range of stepsizes to try out with our agent. Which of the following graphs best represent the range of stepsizes that should be tried out for a given experiment? (the orange points represent the selected stepsizes) 1 point


Average return
per episode
Stepsize


Average return
per episode





12. True or False: Epsilon-greedy exploration uses information from all the action values of a particular state when choosing a *non-greedy* action in that state. 1 point

- (●) False
- ( ) True