Correct
Correct. We can use any state-value learning algorithm.

10. Consider the following state features and parameters θ for three different actions (red, green, and blue): **1 / 1 point**

$$\mathbf{X}(s) = \begin{bmatrix} 0.1 \\ 0.3 \\ 0.6 \end{bmatrix} \qquad \boldsymbol{\theta} = \begin{bmatrix} 45 \\ 73 \\ 21 \\ 120 \\ 120 \\ -10 \\ -100 \\ 200 \\ -25 \end{bmatrix} \begin{matrix} \\ \\ \} a_0 \\ \\ \\ \} a_1 \\ \\ \\ \} a_2 \end{matrix}$$

Compute the action preferences for each of the three different actions using linear function approximation and stacked features for the action preferences.

What is the action preference of a_0 (red)?

- (●) 39
- ( ) 35
- ( ) 37
- ( ) 33

Correct
Correct.

11. Which of the following statements are true about the Actor-Critic algorithm with softmax policies? (**Choose all that apply**) **1 / 1 point**

- [ ] The actor and the critic share the same set of parameters.
- [ ] The preferences must be approximated using linear function approximation.
- [x] The learning rate parameter of the actor and the critic can be different.

Correct
Correct! In practice, it is preferable to have a slower learning rate for the actor so that the critic can accurately critique the policy.

- [x] Since the policy is written as a function of the current state, it is like having a different softmax distribution for each state.

Correct
Correct!

12. Which one is a reasonable parameterization for a Gaussian policy? **0 / 1 point**

- ( ) μ: a linear function of parameters, σ: the exponential of a linear function of parameters.
- ( ) μ: a linear function of parameters, σ: a linear function of parameters
- (●) μ: the exponential of a linear function of parameters, σ: a linear function of parameters.

Incorrect
Incorrect. Remember that the parameter sigma must be positive. A linear function does not guarantee that this constraint will be met.