CS 4700, Foundations of Artificial Intelligence
Spring 2020
Solutions to Quiz 2

1. There were two possible questions, both True/False:

   - If all arms in a multi-armed bandit have been pulled the same number of times then UCB will pick an arm that has the largest cumulative reward.

     True. The general UCB formula for each arm is $\frac{Sum_{\mathrm{i}}}{N_{\mathrm{i}}} + \frac{g(N)}{\sqrt{N_{\mathrm{i}}}}$. Let's say there were n arms and each was pulled k times, so that $N_1 = N_2 = \cdots = N_{\mathrm{n}} = k$. The formula simplifies to $\frac{Sum_{\mathrm{i}}}{k} + \frac{g(nk)}{\sqrt{k}}$. All the terms within this are the same for all arms except for the Sum_{i}, so if we're finding which is the max we want whichever Sum_{i} is largest – the one that has the largest cumulative reward.

   - If all arms in a multi-armed bandit have the same cumulative reward then UCB will pick an arm that has been tried the fewest number of times.

     True. The general UCB formula for each arm is $\frac{Sum_{\mathrm{i}}}{N_{\mathrm{i}}} + \frac{g(N)}{\sqrt{N_{\mathrm{i}}}}$. Let's use S to refer to the common cumulative reward that the arms share, so this simplifies to $\frac{S}{N_{\mathrm{i}}} + \frac{g(N)}{\sqrt{N_{\mathrm{i}}}}$. The only part that depends on N_{i} are the denominators. Note that if $N_{\mathrm{i}} > N_{\mathrm{j}}$ then $\frac{1}{N_{\mathrm{j}}} < \frac{1}{N_{\mathrm{i}}}$ and $\frac{1}{\sqrt{N_{\mathrm{j}}}} < \frac{1}{\sqrt{N_{\mathrm{i}}}}$. These in tern mean that if $N_{\mathrm{i}} > N_{\mathrm{j}}$ then $\frac{S}{N_{\mathrm{j}}} + \frac{g(N)}{\sqrt{N_j}} > \frac{S}{N_{\mathrm{i}}} + \frac{g(N)}{\sqrt{N_{\mathrm{i}}}}$ (regardless of g(N), because it's the same on both sides. In other words whichever one has lower number of pulls (N_{j}) gets a larger UCB value. The one with the lowest number of pulls will be the one with the largest UCB value and will be pulled.

2. There were two possible questions, both True/False:

   - In Monte Carlo Tree Search the branching factor at a leaf node is strictly greater than the number of playouts that have visited the node.

     True. Let's refer to the number of moves that could be applied to a given leaf state as b, and the number of playouts that have gone through this state as a. This means there have been b-a moves that have not been tried yet. To satisfy the definition of a leaf node some of the state's b moves have never been tried in any game playout yet, or in other words b-a>0. This means b>a.

   - In Monte Carlo Tree Search a playout from the root node can fail to reach a leaf node.

     True. If a playout ever hit a terminal node there are no successors and the playout would halt. For example, consider the case where the root was itself a terminal state, or if it was a small enough game and enough playouts had taken place that every possible state had been generated so that there are no longer states that have had moves left unexplored.

3. There were three possible questions, all identical except for the numbers in the question:

   - Consider a multi-armed bandit using the UCB algorithm with $g(N) = \sqrt{2 \ln N}$. Imagine there are two arms. Arm A has been pulled 5 times for a cumulative reward of 10. Arm B has been pulled 4 times for a cumulative reward of 9. What are the UCB values for the two arms? Please give your answer to two significant digits after the decimal: 0.555 should be written 0.56, and 999.994 should be written 999.99.Consider a search problem with branching factor b.

UCB (Arm A) = $\frac{10}{5} + \frac{\sqrt{2\ln 9}}{\sqrt{5}} = 2.94$

UCB(Arm B) = $\frac{10}{4} + \frac{\sqrt{2\ln 9}}{\sqrt{4}} = 3.30$

- Consider a multi-armed bandit using the UCB algorithm with $g(N) = \sqrt{2\ln N}$. Imagine there are two arms. Arm A has been pulled 6 times for a cumulative reward of 10. Arm B has been pulled 5 times for a cumulative reward of 9. What are the UCB values for the two arms? Please give your answer to two significant digits after the decimal: 0.555 should be written 0.56, and 999.994 should be written 999.99.Consider a search problem with branching factor b.

  UCB(Arm A) = $\frac{10}{6} + \frac{\sqrt{2\ln 11}}{\sqrt{6}} = 2.56$

  UCB(Arm B) = $\frac{10}{5} + \frac{\sqrt{2\ln 11}}{\sqrt{5}} = 2.78$

- Consider a multi-armed bandit using the UCB algorithm with $g(N) = \sqrt{2\ln N}$. Imagine there are two arms. Arm A has been pulled 7 times for a cumulative reward of 10. Arm B has been pulled 6 times for a cumulative reward of 9. What are the UCB values for the two arms? Please give your answer to two significant digits after the decimal: 0.555 should be written 0.56, and 999.994 should be written 999.99.Consider a search problem with branching factor b.

  UCB(Arm A) = $\frac{10}{7} + \frac{\sqrt{2\ln 13}}{\sqrt{7}} = 2.28$

  UCB(Arm B) = $\frac{10}{6} + \frac{\sqrt{2\ln 13}}{\sqrt{6}} = 2.42$