Admin

- Our office hours are on the github page
- If you have a question about the assignments, then use eclass, so others benefit from your question
- Bring paper or tablet to class for Friday. We will not print the worksheets---save the trees

Plan for today

- Cover expectations about discussion questions: the good, the bad, the ugly
- Clarify some misunderstandings
- Pose discussion topics
- Break into discussion groups
- Report back to the class about your learnings

Bad Discussion Questions

1. *email Zero*
2. whis is stepsize means? *Spelling, grammar, ?*
3. When will I be able to plug in my brain to download a better OS?
4. For a stationary problem, what is a possible disadva *Incomplete*
5. In RL you are not told which arm is the best. In supervised learning you have the correct answer and the program modifies it's output based on if its estimate of the correct choice is correct or not. *No question or topic of discussion*
6. Will we have access to the lecture slides notes prior to the actual lecture? *Admin, ask on eclass*
7. The second question of the quiz, I tried to plug the coordinate into the formula, however I can only get 1/7 which closes to 1/8, and I cannot get constant stepsize in that question. *Ask on eclass or go to office hours for help*
8. In Assignment 1, what's thew different between updating self.q_values first and updating current_action first? *Ask on eclass or go to office hours for help*
9. how to identify the estimat is updated by the prediction error?(ex. 1/2 or 1/(t-1))? *Ask on eclass or go to office hours for help*

Review & clarifications

- Is the action which has highest expected reward(value) the optimal action?

- What exactly the stepsize is? The meaning and influence of it? and why it could be constant?

- Why does epsilon of a greedy agent perform worse than some epsilon? Does greedy agent always choose the best action?

- Why setting optmistic initial values is not well suited for non-stationary problems?

- Is it possible that in e-greedy, with probability e, the action taken by agent randomly could be the greedy action again?

Review & clarifications (2)

- Is it possible to combine e-greedy method with optimistic initial values?

- In what conditions the constant step size will be the most useful?

- how do we include a price for taking actions?

- The update rule we use has an issue of vanishing weights (old data is less important to decisions), what are the benefits of this and would we maybe be better served by using something like a rolling average?

Review & clarifications (3)

- What would be the best way to change what we've learned so far if the goal is not to arrive at the best option in the long term, but instead to achieve a certain minimum as fast as possible?

- How can solutions to the k-armed bandit problem be modified to handle cases where the decision you make now could potentially change the possible rewards that you could achieve from each arm in the future? (delayed outcome)

    - or if the sequence of actions matters, or some other variant ...

- What does convergence mean?

Discussion topics for today

1. **How do we set** c, epsilon, alpha, or the initial values? Can the agent do it themselves from data?

2. Discuss some **real-life applications** that can be modelled as a **k-armed bandit problem**

3. What **other approaches** could we use to solve the k-armed bandit problem?

4. How and when do our agent's **stop exploring**?

5. Can we design agents that are **robust to non-stationarity**? Can we build agents that notice when the world changes? If yes what could they do about it?

6. When is RL a **bad fit** for an application

7. Humans use ***intuition*** to deal with new situations. What is intuition? Can machines have intuition?

8. ***D***oes the distribution of rewards matter? Do the algorithms leverage the fact that the rewards were sampled from a Gaussian?