Reinforcement Learning: Chapter 2 Notes

Name: Eli Andrew

- Most important feature distinguishing reinforcement learning from other types of learning is that it uses training information that *evaluates* the actions taken rather than *instructs* by giving correct actions.

- In our k-armed bandit problem, each of the k actions has an expected or mean reward given that that action is selected; let us call this the *value* of that action. We done the action selected on time step t as A_t, and the corresponding reward as R_t. The value then of an arbitrary action a, denoted $q_*(a)$, is the expected reward given that a is selected: $q_*(a) = E[R_t|A_t = a]$ The estimated value of action a at time step t is denoted as $Q_t(a)$. We would like $Q_t(a)$ to be close to $q_*(a)$.