Reinforcement Learning: Chapter 1 Notes

Name: Eli Andrew

- The distinction between problems and solution methods is very important in reinforcement learning
- While reinforcement learning seems like a kind of unsupervised learning because it is not relying on examples of correct behavior, it is actually trying to maximize a reward signal instead of trying to find hidden structure.
- Methods like genetic algorithms never estimate value functions. They apply multiple static policies each interacting over an extended period of time with a separate isntance of the environment. The policies that obtain the most reward, and random variations of them, are carried over to the next generation of policies, and the process repeats.
- Evolutionary methods are effective in small policy search spaces and have advantages on problems in which the learning agent cannot sense the complete state of its enviornment.
- Temporal difference learning is named for changes based on difference $V(S_{t+1}) - V(S_t)$ between estimates at two successive times.
- Evolutionary methods evaluate polcies by holding a given policy fixed and playing many games. The frequency of wins gives an unbiased estimate of the probability of winning with that policy, and can be used to direct the next policy selection.
- Evolutionary methods only change policy after serveral games and only the final outcome of each game is used: what happens *during* the games is ignored. For example, if the player wins, then all of its behavior is given credit, independently of how specific moves might have been critical to the win. This can give credit to moves that never even occured.
- Value methods, in contrast to evolutionary methods, allow individual states to be evaluated. In the end, evolutionary and value function methods both search the space of policies, but learning a value function takes advantage of information available during the course of play.