Reinforcement Learning

1. RL problem statement. State, Action, Reward, Environment, Action
2. Crossentropy method
3. Value function, Q-function
4. Q-learning, approximate Q-learning. DQN, bells and whistles (Experience replay, Double DQN, autocorrelation problem)
5. Policy gradient and REINFORCE algorithm
   a. Baseline idea, A2C
6. Policy gradient applications in other domains (outside RL). How Self-Critical Sequence Training is performed? What is used as a baseline?