Problems

- **Q-learning with value function approximation can be unstable and diverge (do not converges to the optimal q*(s,a))**
  - **P1) highly correlated state**
    - In a continuous environment, the states at t, t+1, t+2, … are almost same
    - It makes generalization difficult
  - **P2) Target oscillation problem**
    - $R + \gamma \max_{a'} q(s', a') - q(s, a)$
    - if the learning model $q(s, a)$ is updated, the learning target $R + \gamma \max_{a'} q(s', a')$ is changed
    - Thus, even though the model has not adequately learned, the target value (true value) change