Reinforcement Learning

- Q - Learning
  - Off-Policy Temporal-Difference Control
    - differentiate behavior-policy from learning-policy
    - SARSA (on policy)
      - *<s, a, r, s', a' ← π(s)>* => Learning
    - Q-Learning
      - *<s, a ← π(s), r, s'>* => Learning
    - Update rule

$$Q(S_t, A_t)_{new} = Q(S_t, A_t)_{old} + \alpha[R_{t+1} + \gamma max_{a \in A} Q(S_{t+1}, a') - Q(S_t, A_t)_{old}]$$

Reinforcement Learning

▪ Function Approximation

- Why function approximation?
  - Problem with large state spaces
    - Large memory for large table task
    - data should be accurate
  - Generalization
    - to generalize from previous encounters with different states that are in some sense similar to the current ones

- Generalization => function approximation
  - to generalize desired functions (e.g value function, q function etc.)
  - utilize *supervised learning*