Reinforcement Learning

- Deep-Q Network
  - Approximate Q(s, a) (Q-function) with deep-neural-network
  - Off-policy learning
    - update rule s equal to q-learning and error is defined by
      - $MSE = (R_{t+1} + \gamma max_{a'} Q(s', a', \theta) - Q(s, a, \theta))^2$
  - using Target network
    - off policy effect: differentiate **learning-policy** from **behave-policy**
    - by *t* time interval, update target network
  - Replay memory
    - fixed-size-queue where store {<*s, a*← *π(s), r, s'*>, <*s, a*← *π(s), r, s'*> ..}
    - to do batch-learning with random sampling
    - stable learning and diverse states are considered to learn

Reinforcement Learning

- Deep-Q Network