

Reinforcement Learning

Value Function Methods

discrete action-space

Discrete state-space

continuous state-space

Dynamic Programming
(model based)

Bellman Equation

DQN

Double Q-Learning
Duelling networks
adding noise
prioritized replay
delayed Q-learning
multi-step Q-learning

Model free

Temporal Difference Learning

SARSA

TD(n)

Expected SARSA

TD(lambda)

Q(sigma)

Monte-Carlo estimates

Actor Critic

A2C, A3C

DDPG - Family

deterministic policy

Dueling DDPG

Prioritized DDPG

TD3

SAC stochastic policy (with sigma as output)

Policy Gradient Methods

continuous action-space

REINFORCE
(Vanilla Policy Gradient)

TRPO
(some degree off-policy)

PPO
(some degree off-policy)
(simplification of TRPO)

on-policy

off-policy

Importance Sampling

Experience Replay

Bootstrapping

Watkins Q-Learning
(Tabular Q)