ML Week 11 – Reinforcement learning

RF learning is loosely based on how humans learning

RL Agent learns by interacting w/ environment

↳ to achieve given goal(s)

RL environment

State → Agent observe the env state/ conditions

$S_t \in S$ @ time stamp t

State has a markov property. dependant on state @ tag of t's . e.g (t), (t, t-1) (t, t-1, t-2)

Action → Agent can select an action @ time t → a_t

→ availble given the state

→ $A(S_t)$

$a_t \in A(S_t)$

reward → the reward that the env gives at t + 1 dependant on the action taken @ t

Arrive @ new state S_{t+1}

▶ RL Agent elements

Agent → · continually interacts w/ timesteps
· obj is to max reward

Policy → Function that specifies the Behaviours available to an agent given the env

State to action

function may be either deterministic or stochastic

value → agents estimation of future reward

▶ reward vs value

reward is short-term gain from near

value long-term total gain

↳ more important to learning

Supervised vs Reinforcement

SL	RL
• Model (weights)	• Agent (Policy)
• input data	• Env state
• Prediction	• Agent Action, next state
• loss func	• Reward
Model is train on existing ground truth (input data)	Agent & Env interact to learn
	Agent learns from own experence (no exist truth)
min loss on train data, generalize to unseen	Max reward via Exploration of actions

Exploit vs Exploration

Exploitation (greedy) → Select highest reward for next action, requires knowledge of action & rewards

Exploration → select action w/ uncertain value to obtain better value estimate

test & learn for future