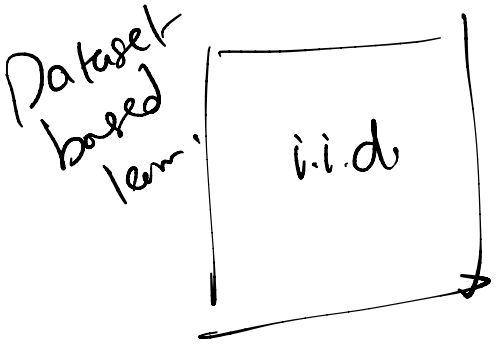
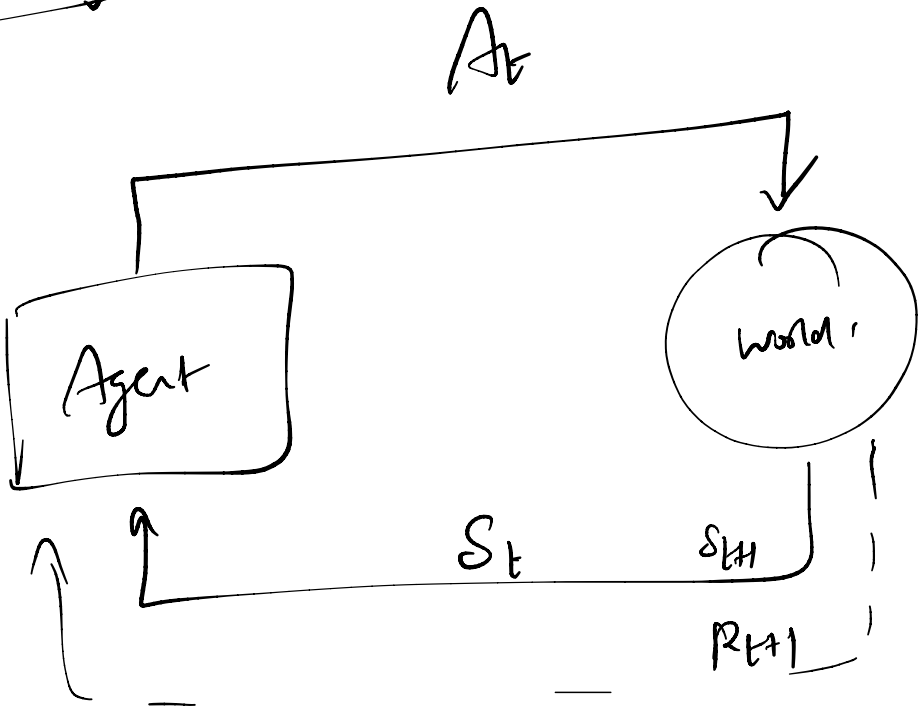
Lecture-13

Reinforcement Learning

→ learning through interaction.





Maximize the return.

$R_t + R_{t+1} - - - -$

→ episodic tasks

→ Continuing tasks. $R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \ldots$

Markov Decision Process: MDP

S, A, T, R, γ

→ State value fn. $V(s)$

→ state-action value fn $Q(s, a)$

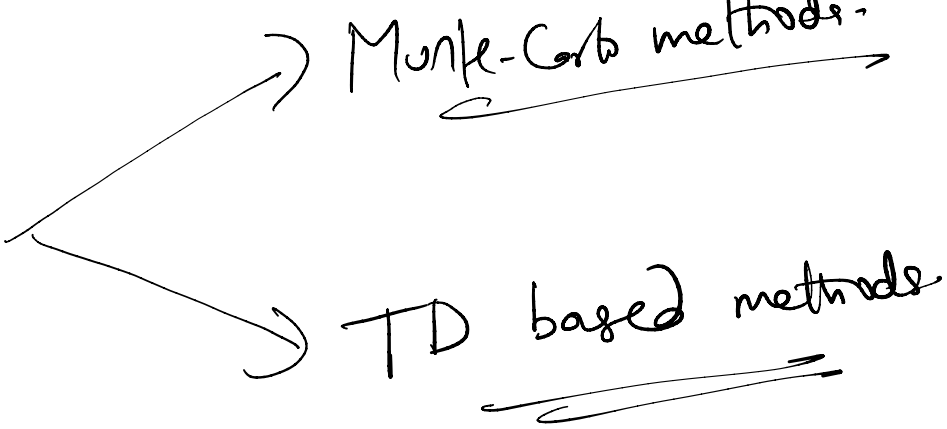
→ Policy $\pi(s)$

→ Value-based

→ Policy gradient methods.

Dynamic Programming

- Bellman equations.
  - → Policy iteration
  - → value iteration.

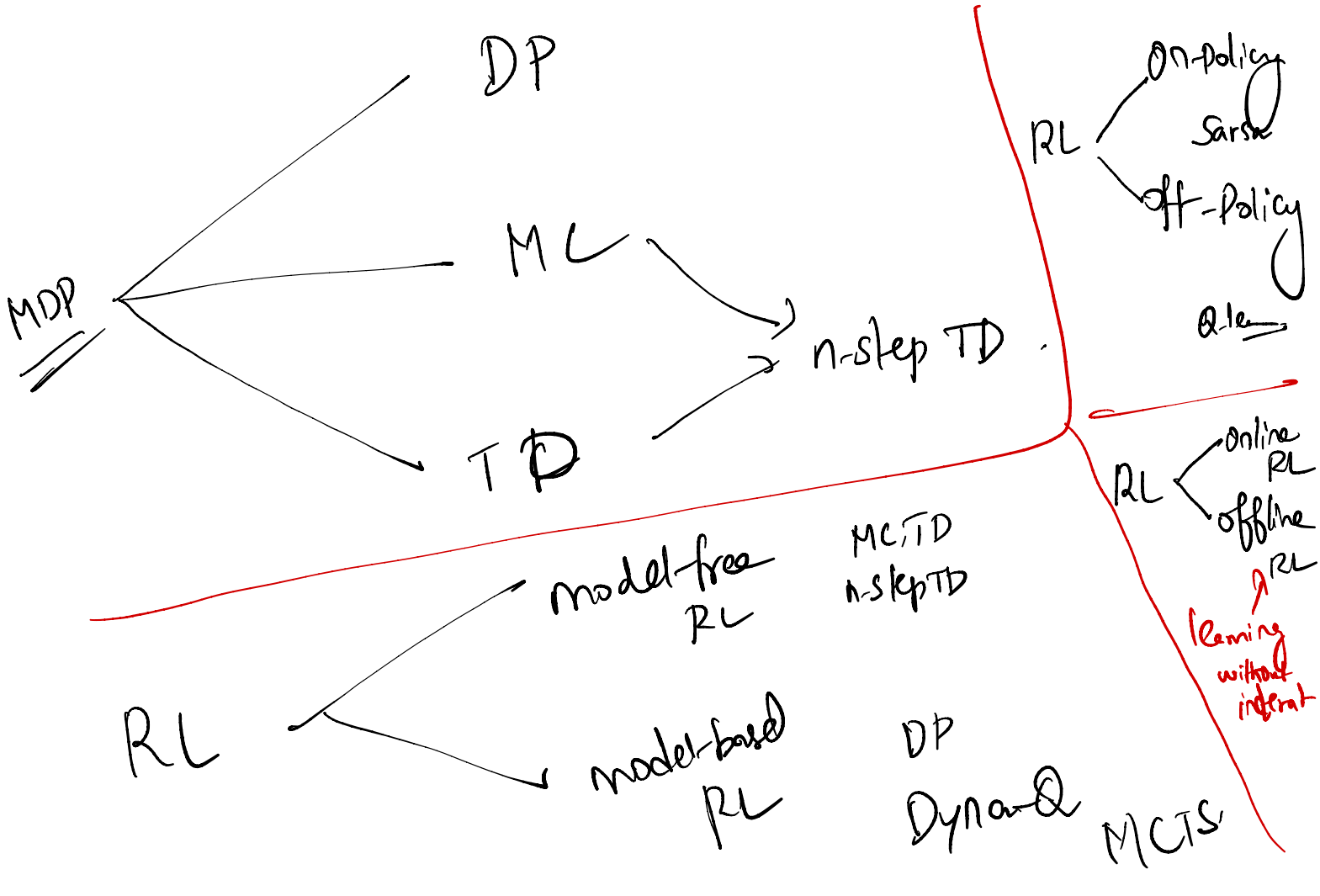


MC

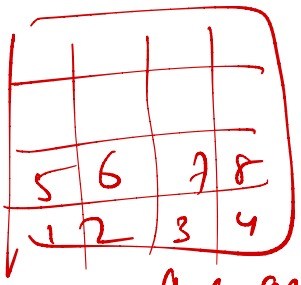
TD

- No bootstrapping.

- bootstrapping.

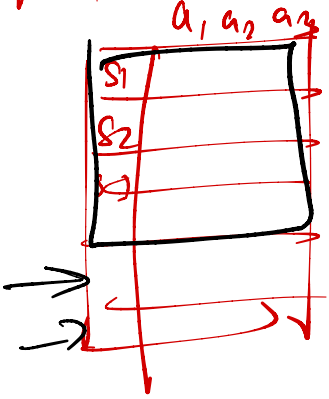


Challenges in Solving RL :-



① Representation learning for RL

→ FA — Scale / Generalize



① Pretraining → CNN → Q

Represent Q-value, δ V(s)

- forward prediction

S_t → CNN → Decon NN → S_{t+1}

a_t

S_t → S_{t+1}

a_t

② Auxiliary tasks

S_t → CNN → A_t → Q

A_t → Decon. → S_{t+1}

② Partial observability

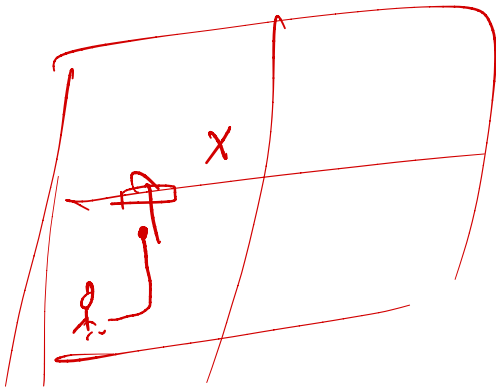
→ POMDP

→ RNNs.

Observations

States

Decision making under uncertainty



③ State and action abstractions

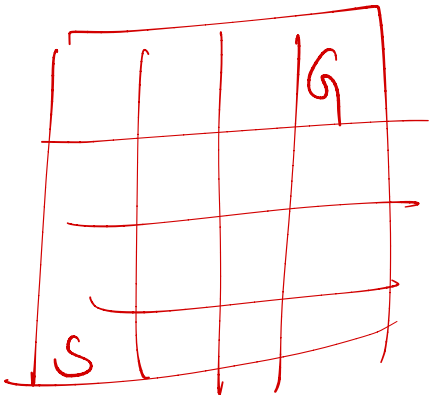
Hierarchical RL — options

— MaxQ.

④ Generalization in RL

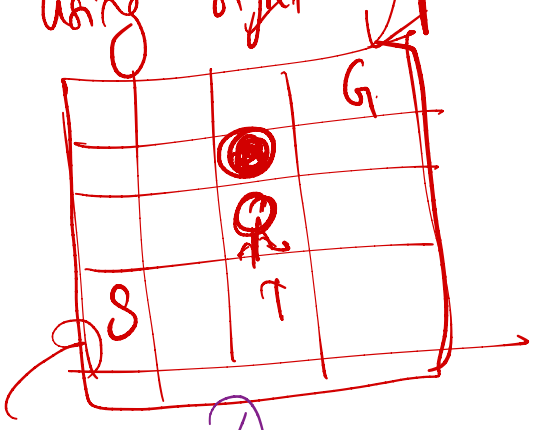
$f: x \rightarrow y$ | Tr

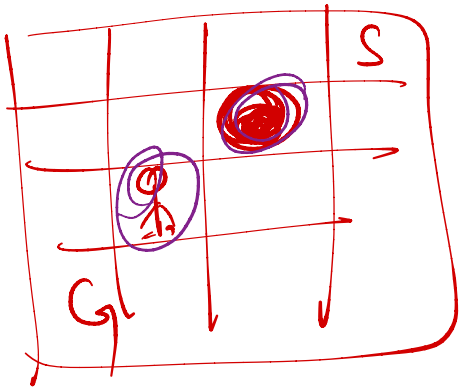
- Generalizat to new States ✓
- Generalize to new tasks. ✓
- Non-stationary env.

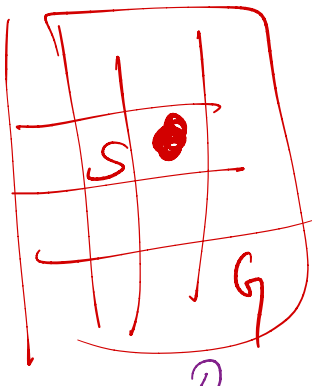


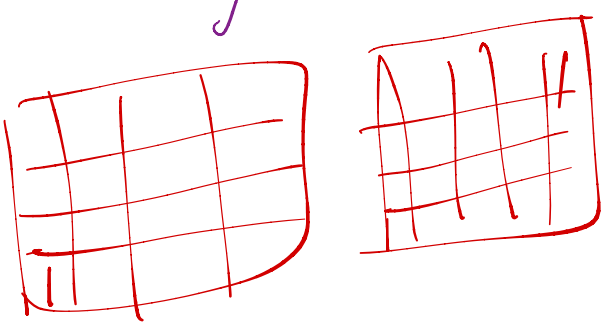
Task: MDP ~ Distribution of MDPs

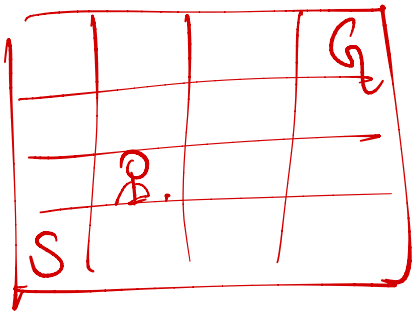
① using right representation.



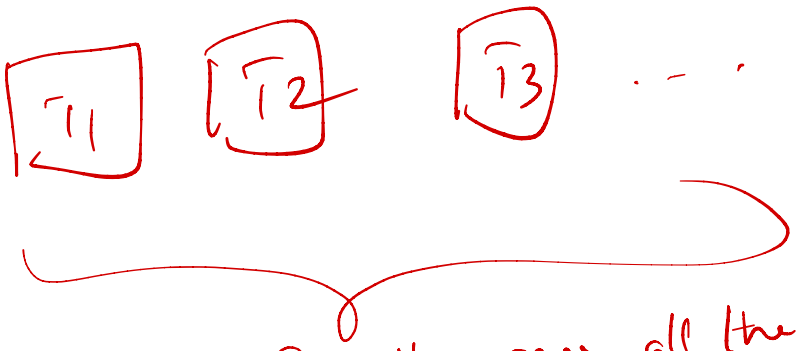






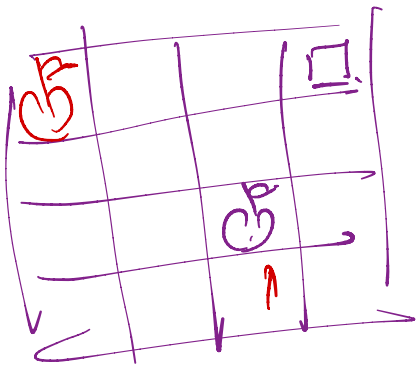


① → Multi-task RL



Simult- access all the tasks.

→ task interference



50%

Apple → +10

box → -10

50%

Apple → +10

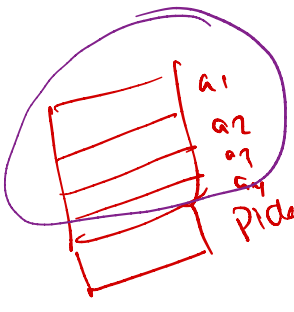
box → +10

② → lifelong RL / continual RL

→ Sequence of tasks.

① knowledge transfer
- Forward transfer.
- Backward transfer

→ hierarchy / abstraction



→ Self-Supervised learning

Catastrophic forgetting:

→ Continual learning (Udem - next winter)

→ RL Course at Udem (Fall)

→ Deep learning (Udem → Fall/Winter)

→ RL Course → DeepMind (overlap)

→ DeepRL Course — Sergey Levine UCB
