Week 1 – 2 – Studying & Researching

Week 3 – 5 – Implementing Hierarchical learning

Week 6 – 8 – Implementing Alpha star and general improvements

Week (spring break) – Writing Paper and Short video presentation

Week 9 - Rosie Submission

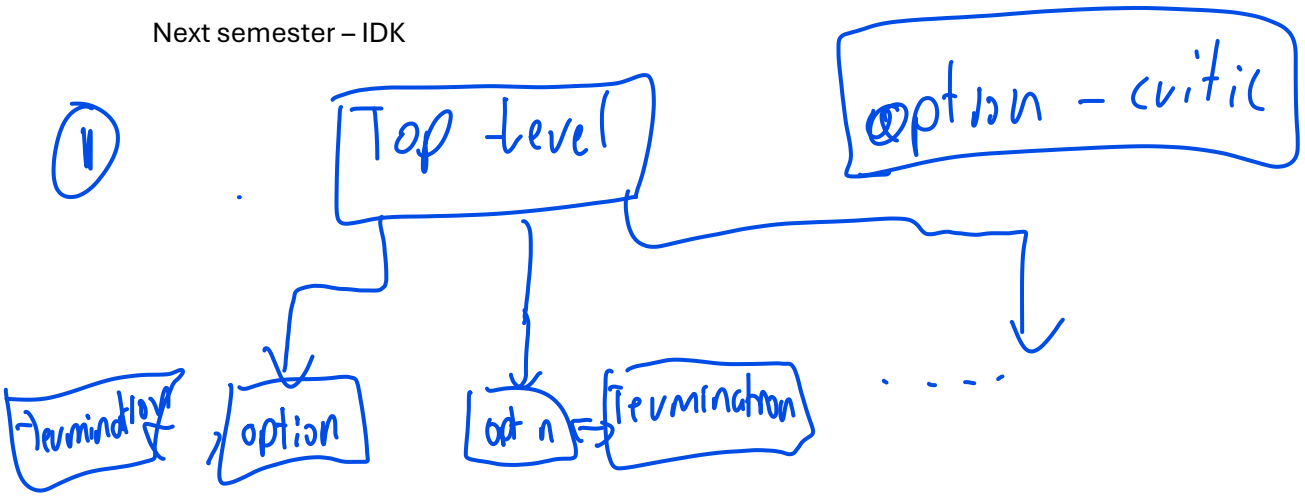
Week 9 - 12 – Continued improvements and finalized results (for Rosie)

Week 11 – 13 – Presentation preparation

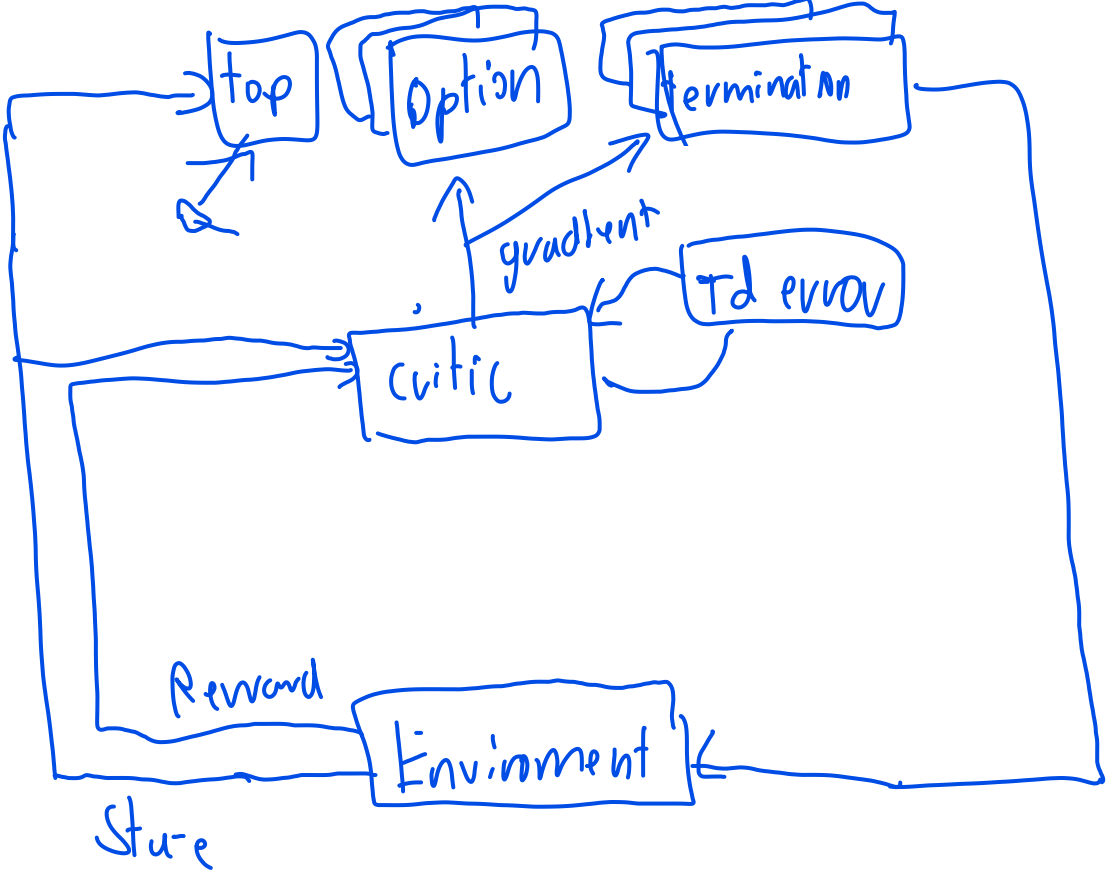
Week 13 – Rosie competition

Week 14 – 16 – Continue work and work on requirements of UR

Next semester – IDK




- Top level - decides option
- Option - decide action
- Termination - decides when to stop option

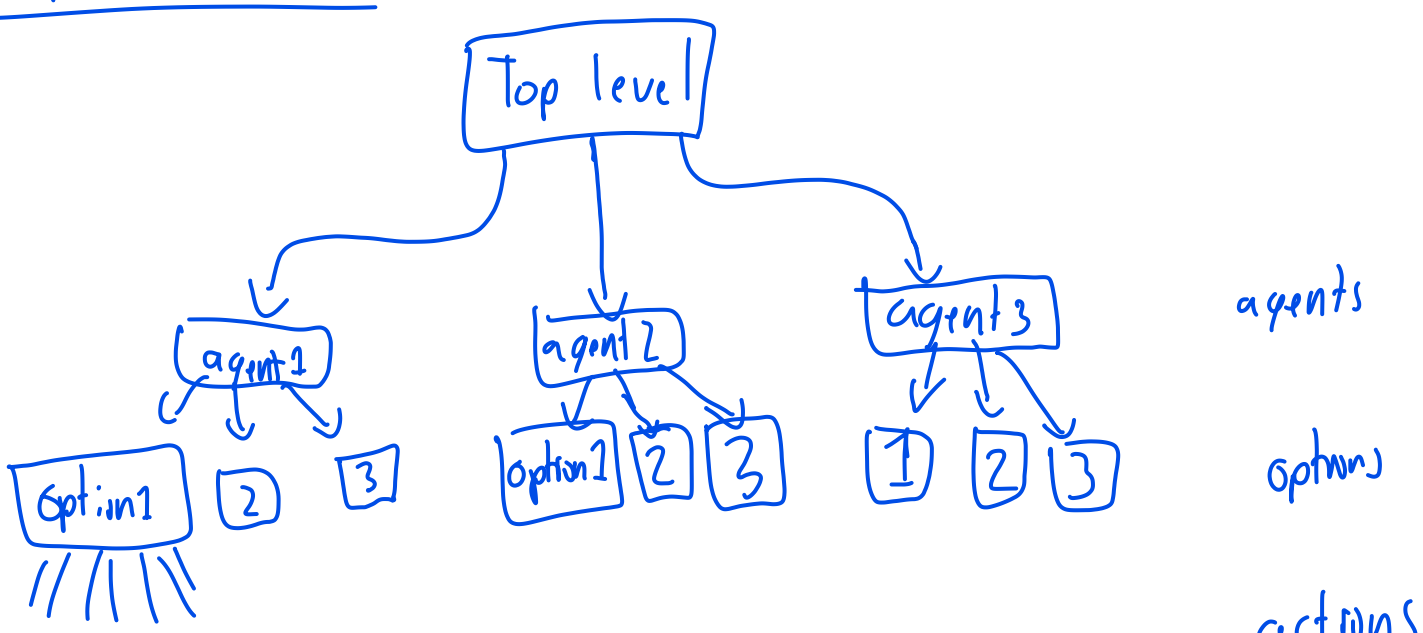



---

(2) DIAYN - promotes diversity not useful res)

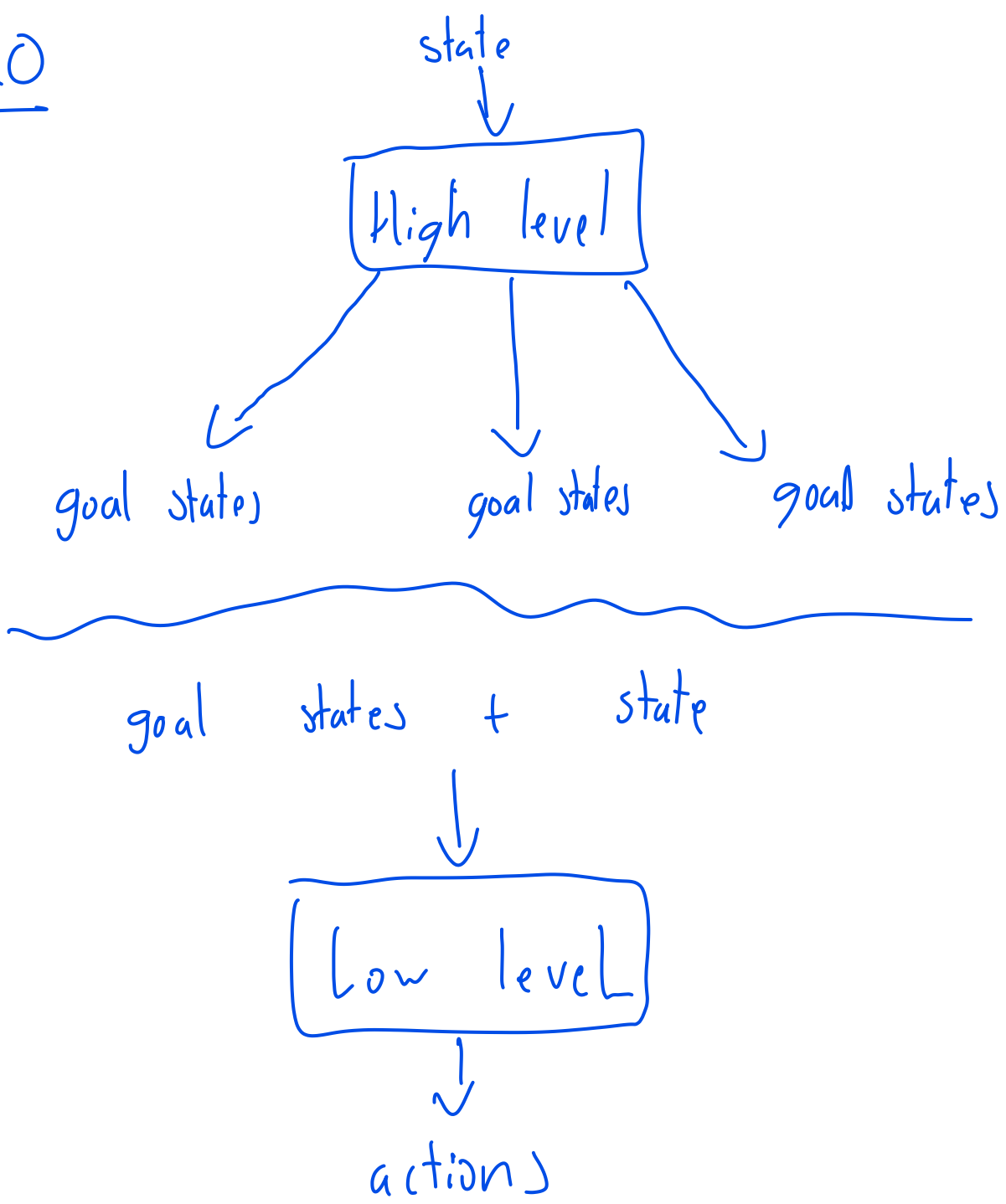
- discriminat - Identifies option based on action & state
- agent - rewarded if discriminator is right

# Option-Critic




- everything trained through advantage of critic
- regularization at top level can be used
- DIAYN - Diversity: discriminator can be used to promote diversity of different options

# HIRO




- Low level trained $\sqrt{(\text{goal state} - \text{state})^2}$
- Top level trained reward function / critic

$$r = -\|s' - (s - g)\|^2$$

# Hiro pt 2

- Low level policies could become overly goal focused?
- Opponent position should not be accounted for?
- reward Low-Level policies with advantage
- reward top-Level with enemy prediction
- reward ball_state matching move with agent_state?
  goal ball state goal agent state

- Value iteration
  - Q table : blank
  - Policy : greedy selection of Q table
  - updates value after each episode

$$V(s) \leftarrow V(s) + \text{Reward} + \gamma((\text{Probs})\, V(s'))$$

  - till convergence

- Policy iteration
  - Policy randomly selected and fixed
  - Value iteration on fixed policy till convergence
  - new Fixed policy based on greedy selection on old policy value estimate
  - Same update rule