TD3 for Competitive Air Hockey

Reinforcement Learning Final Project

Julian Jurcevic

RL Course WS 2025/26

Problem Setting

Environment

- Competitive 2D Air Hockey (Gymnasium + Box2D)
- Continuous control (translation, rotation, shooting)
- Zero-sum two-player game

Core Challenges

- Continuous high-dimensional action space
- Function approximation error
- Non-stationarity due to opponent

TD3: Core Idea

Clipped Double Q-Learning

$$y = r + \gamma(1 - d) \min_{i=1,2} Q_{\phi'_i}(s', \tilde{a}')$$

Target Policy Smoothing

$$\tilde{a}' = \text{clip}\big(\pi_{\theta'}(s') + \epsilon\big)$$

Delayed Actor Updates + Polyak Averaging

Replay & Exploration

Off-policy learning

- Replay buffer
- Uniform + Prioritized Experience Replay

Noise Annealing

$$\sigma_t = \max\left(\sigma_0(1 - t/T), \sigma_{\min}\right)$$

Large exploration early — stable exploitation later

Curriculum & Self-Play

Stage I

- Weak opponent only

Stage II

- Mixed curriculum (weak + strong)

Stage III

- Increased strong + self-play

Improves robustness and prevents overfitting.

Final Results

Variant	WR Weak	WR Strong
Scratch	61%	35%
Pretrained	79%	52%
+ PER	82%	59%
+ Self-Play	**89%**	**71%**

Best performance: Pretraining + PER + Self-play

Conclusion

- TD3 is effective for competitive continuous control
- Curriculum stabilizes learning
- Self-play significantly improves performance
- Final agent wins consistently against strong baseline

Thank you.