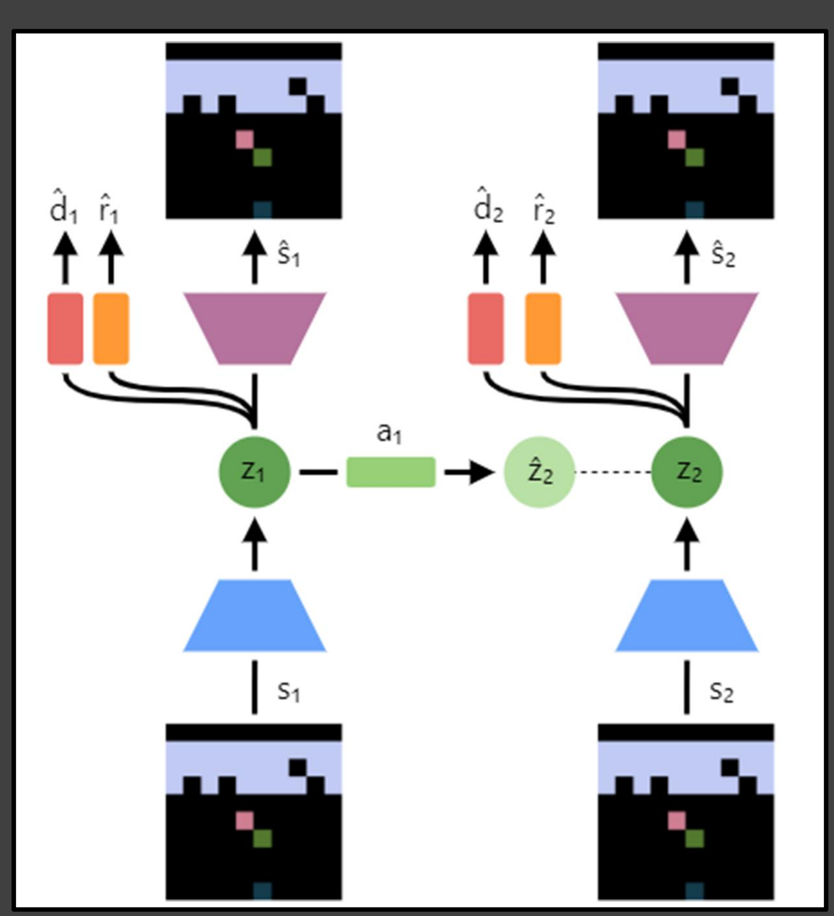
Acting in the Face of Uncertainty
Pessimism in Model-Based Reinforcement Learning

1. Introduction

- Model-Based RL (MBRL) creates a model of the environment to generate policies.
- Challenges in MBRL: inaccurate world models in underexplored regions of the training set can lead to suboptimal decisions.
- Pessimism aims to address this issue by disincentivizing actions leading to out-of-distribution (OOD) states.
- **Research question: How does incorporating pessimism in the planning loop affect agent performance?**

2. Model Architecture

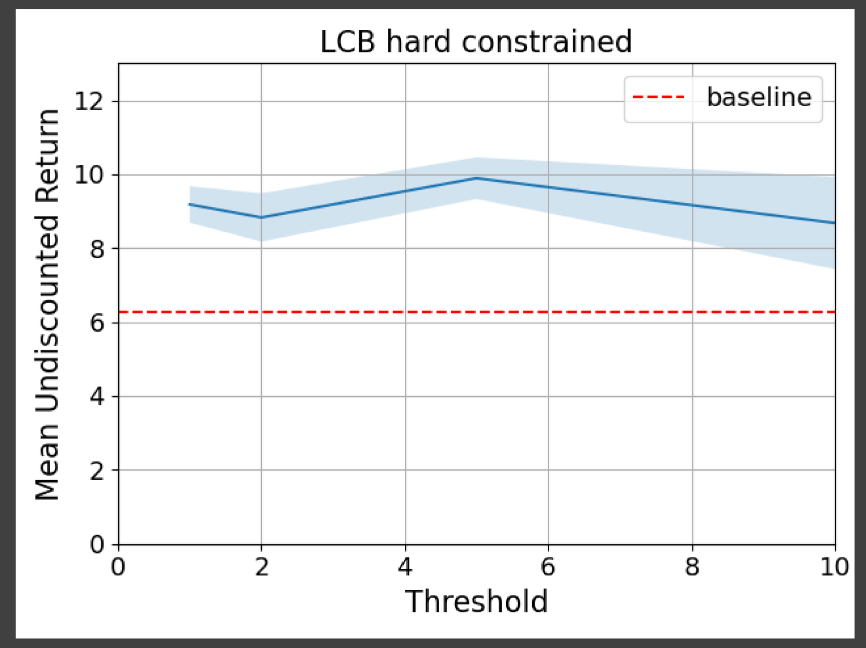


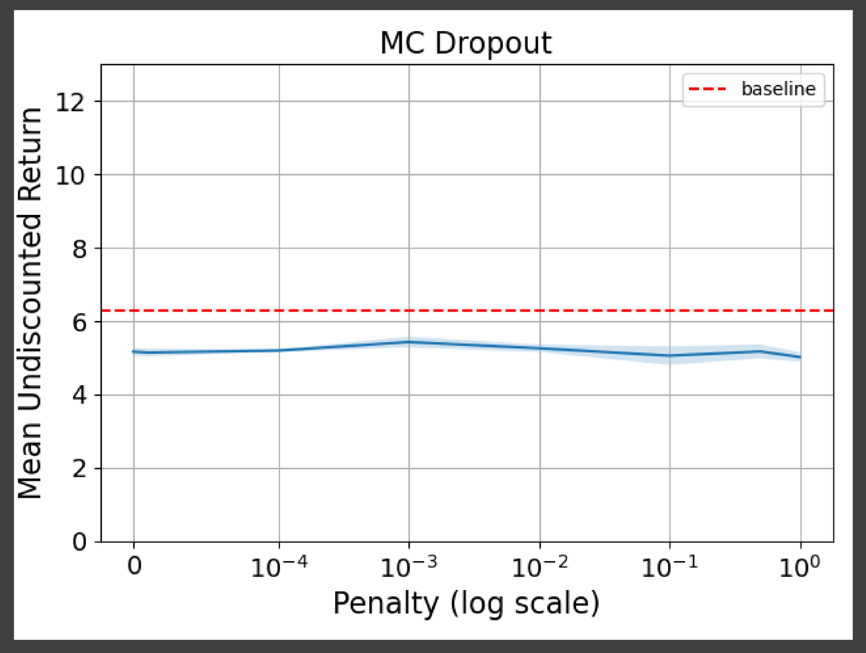
- The world model uses a representation, dynamics and reward network to plan.
- The agent uses a Monte Carlo tree search with the world model to chose actions.

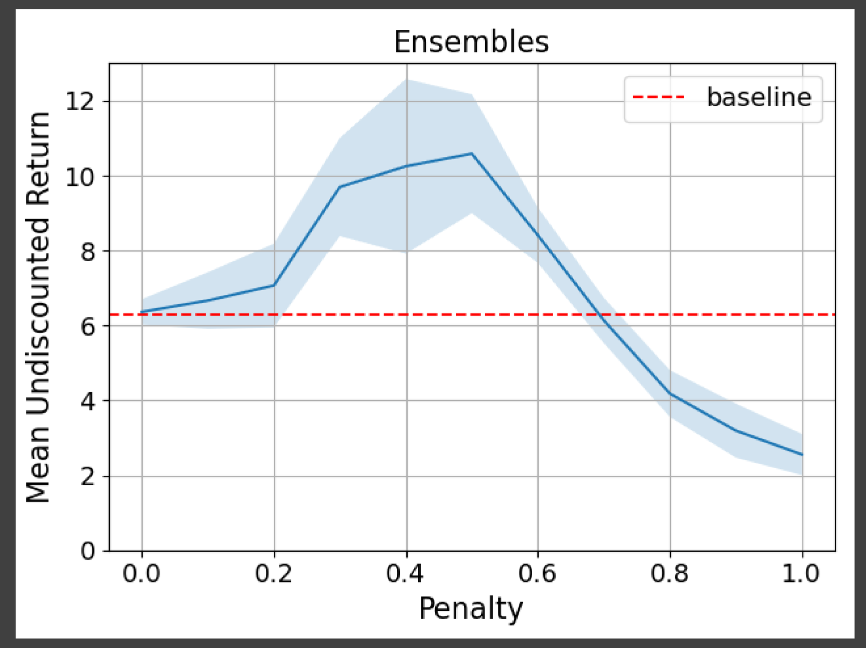
3. Pessimism Approaches

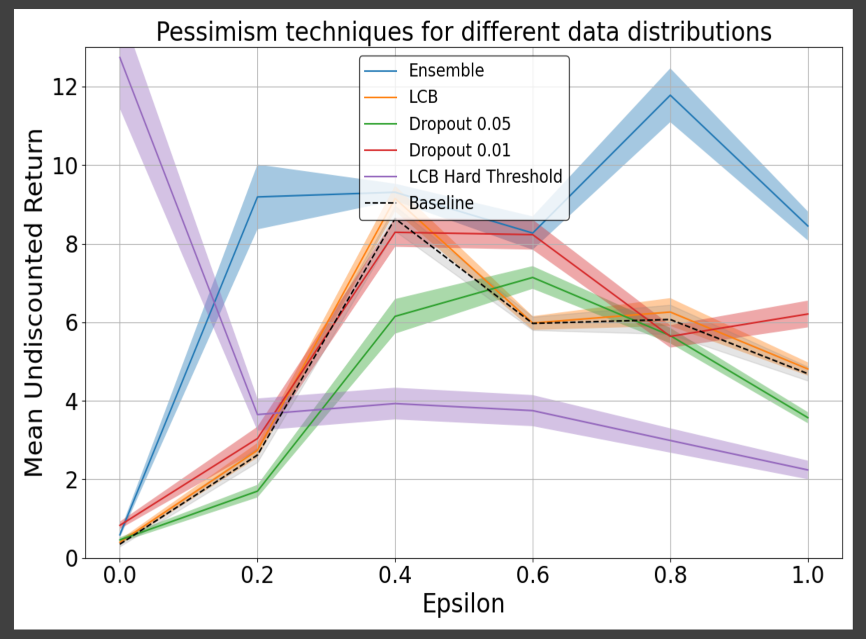
- Pessimism requires the quantification of the uncertainty in the world models predictions.
- Investigated Techniques: Lower Confidence Bound (LCB), Ensembles, Monte Carlo (MC) Dropout
- **LCB [1]** penalizes actions based on the counts of state-action pairs in the training set. A penalty can be enacted, or the planning can be constrained to only actions above a threshold.
- **Ensembles [2]** combine the results of multiple diverse neural networks to quantify uncertainty.
- **MC Dropout [3]** uses dropout at model inference time to perform stochastic forward passes from which uncertainty can be quantified.

4. Results









5. Conclusion

- Ensembles provide the highest performance gains over the widest range of data distributions.
- LCB shows variable improvement. It is only effective with near-optimal datasets
- MC Dropout generally not effective, but it may benefit from increased model size.

6. Future Work

- Apply methods in more complex environments.
- Methods for automatic penalty coefficient optimisation.
- Investigate MC dropout with larger models or uncertainty decomposition.

Sten van Wolfswinkel | s.k.vanwolfswinkel@student.tudelft.nl
Supervisor: Jinke He | Responsible Professor: Frans Oliehoek

References

[1] Rashidinejad, P., Zhu, B., Ma, C., Jiao, J., & Russell, S. (2023). *Bridging Offline Reinforcement Learning and Imitation Learning: A Tale of Pessimism* http://arxiv.org/abs/2103.12021

[2] An, G., Moon, S., Kim, J.-H., & Song, H. O. (2021). *Uncertainty-Based Offline Reinforcement Learning with Diversified Q-Ensemble* https://doi.org/10.48550/arXiv.2110.01548

[3] Wu, Y., Zhai, S., Srivastava, N., Susskind, J., Zhang, J., Salakhutdinov, R., & Goh, H. (2021). *Uncertainty Weighted Actor-Critic for Offline Reinforcement Learning* https://doi.org/10.48550/arXiv.2105.08140