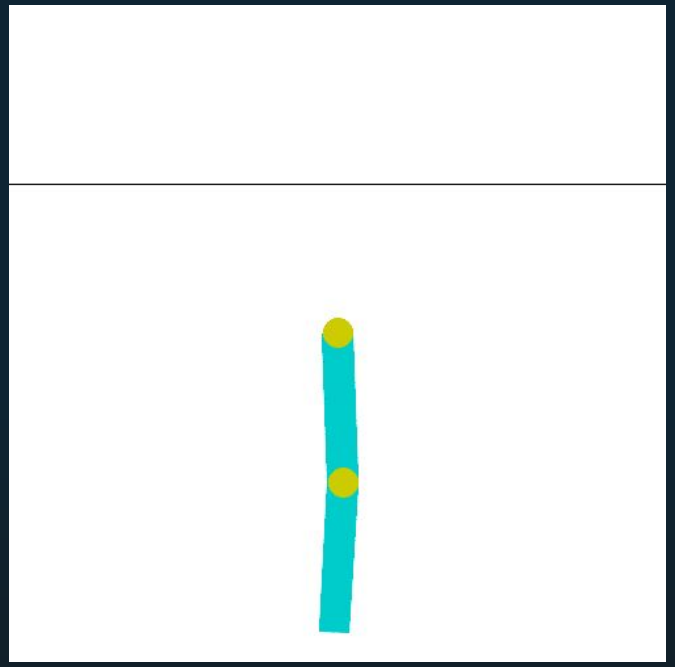
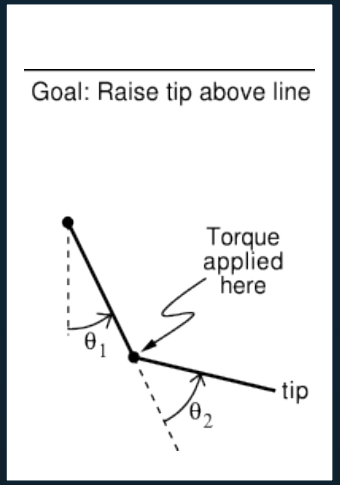
Range-to-Range Regression

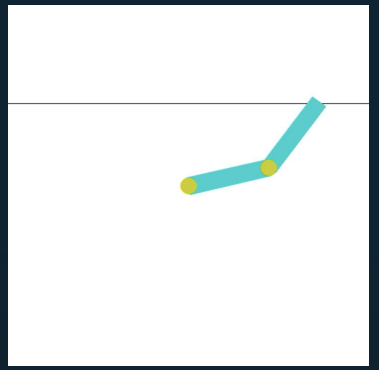
Propagating Uncertainty in Environment Model

Zoe Shao, L.A.C.E. Lab
Advisor: Erin Talvitie

Reinforcement Learning



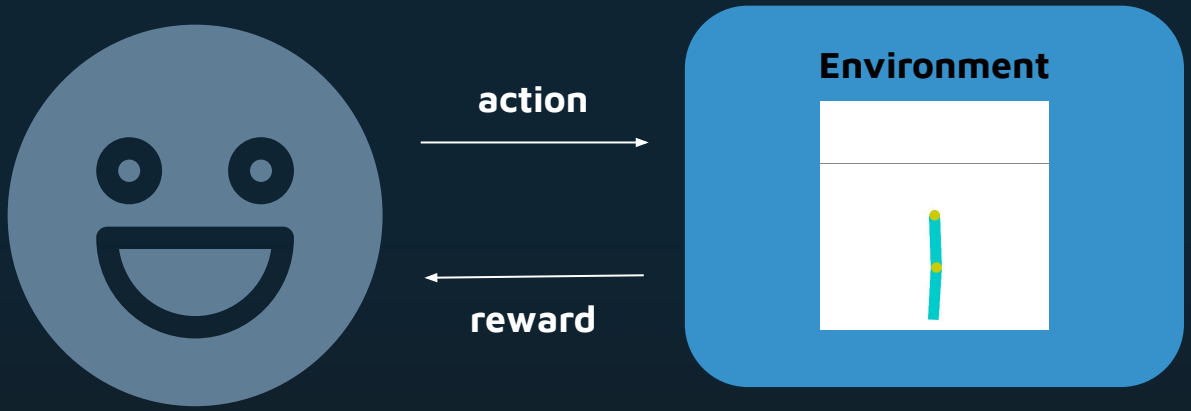




Goal: reach the top in fewest frames

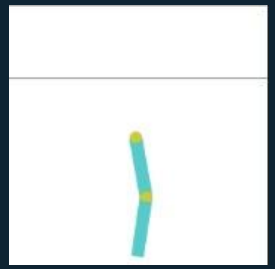
Source: [Gym Documentation - Acrobot]

What is Reinforcement Learning (RL)?

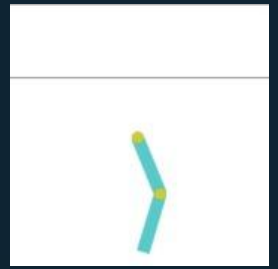


How does the agent learn?

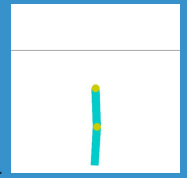
Sample From Real Environment



Torque 1



Simulated Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

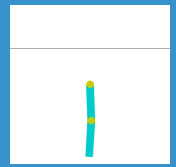
Est.
Torque 1: 80
Torque 0: 90
Torque -1:100

Target: 1+80 = 81

action

reward

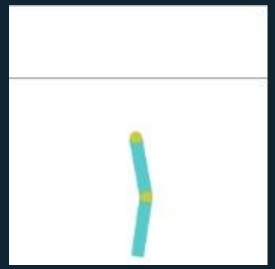
Environment



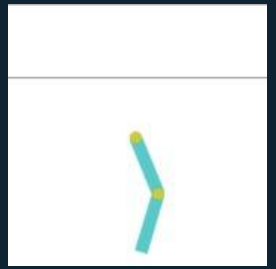
* Est. is the estimate number of frames from the goal

How does the agent learn?

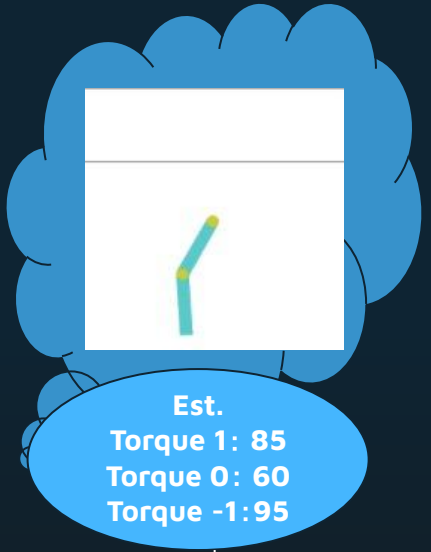
Sample From Real Environment



Torque 1



Torque 1



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Est.
Torque 1: 80
Torque 0: 90
Torque -1:100

Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

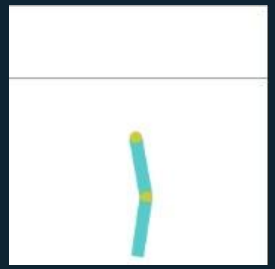
Target: 1+80 = 81

Target: 2+60 = 62

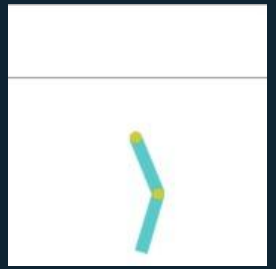
* Est. is the estimate number of frames from the goal

Planning

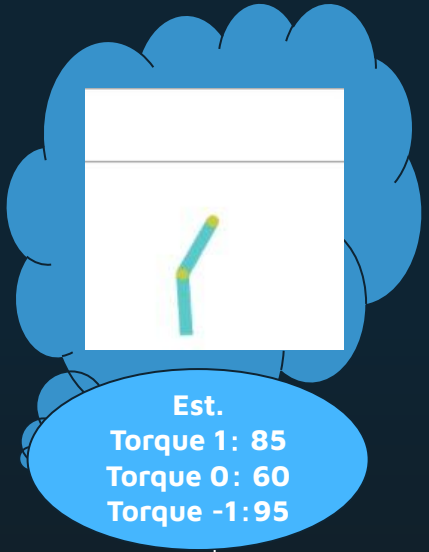
Sample From Real Environment



Torque 1



Torque 1



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Est.
Torque 1: 80
Torque 0: 90
Torque -1:100

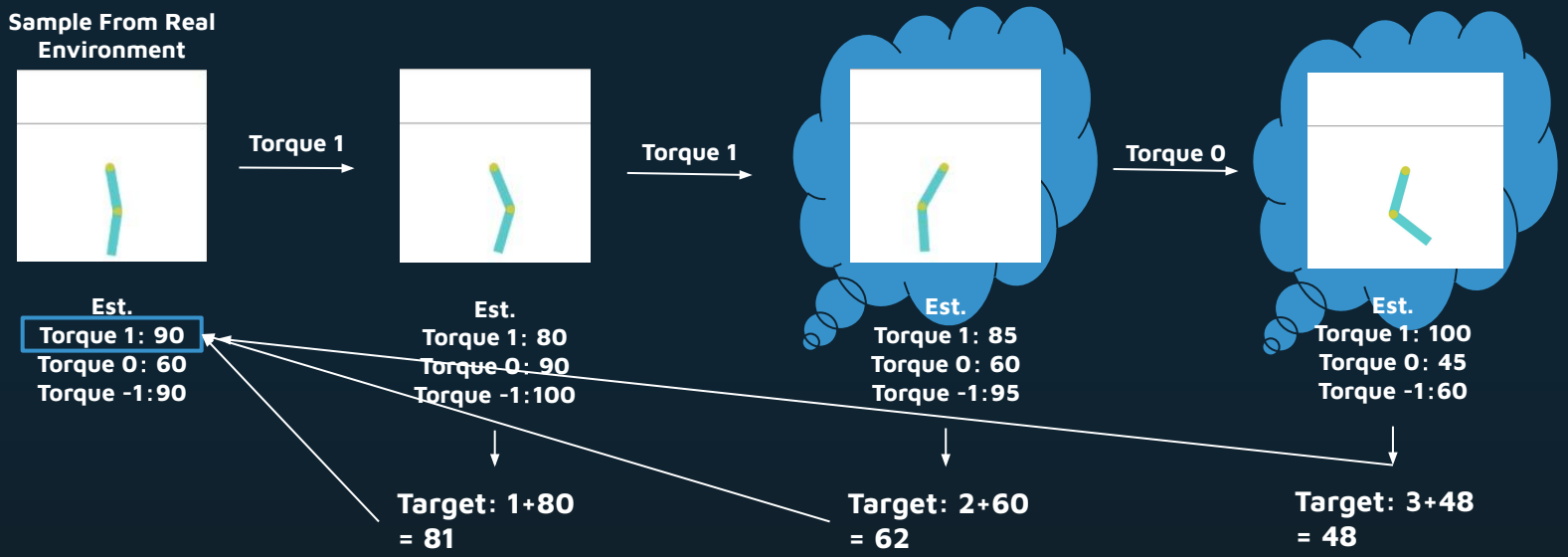
Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

Target: 1+80 = 81

Target: 2+60 = 62

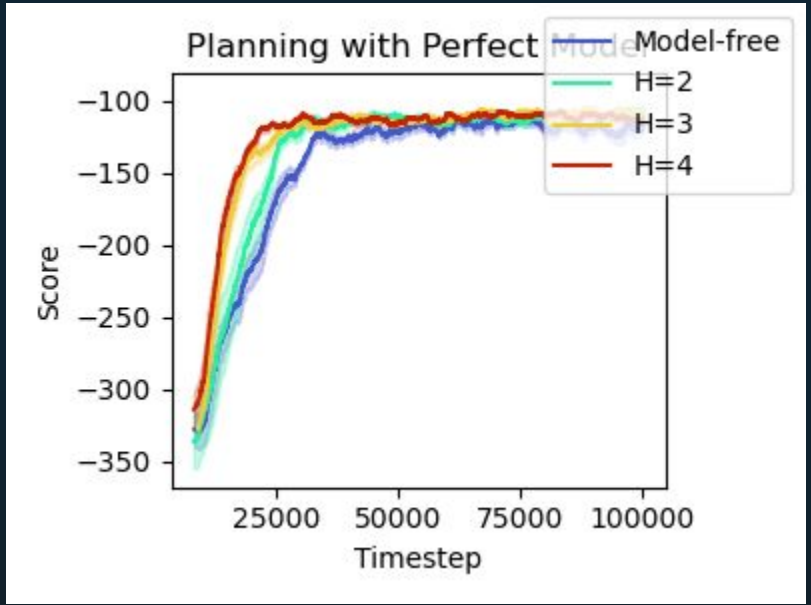
* Est. is the estimate number of frames from the goal

3-Step Planning



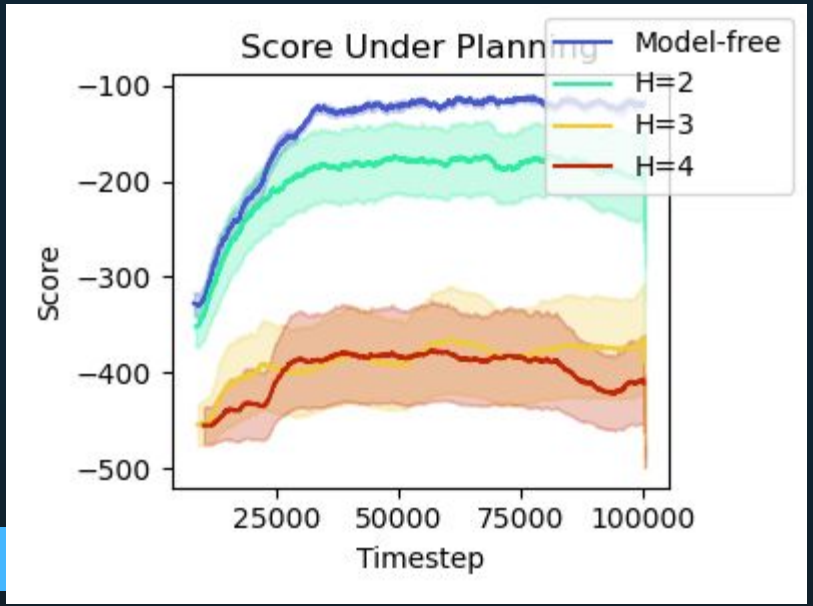
* Est. is the estimate number of frames from the goal

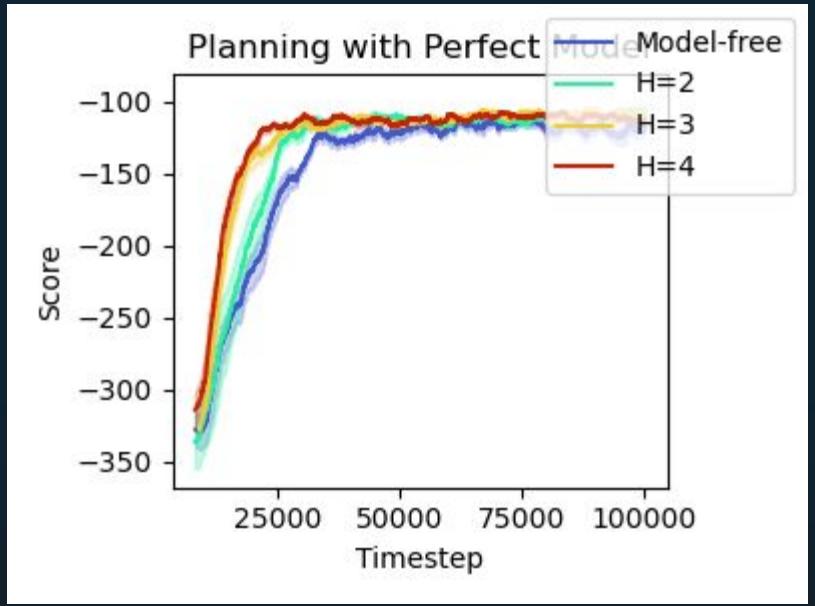
Planning Helps



Planning with Perfect Model

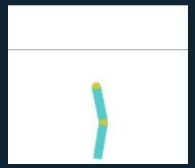
Planning Not Helping?





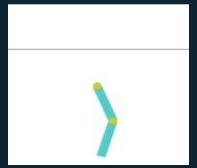
Planning Not Helping?

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Torque 1



Est.
Torque 1: 80
Torque 0:90
Torque -1:100

Torque 1

Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

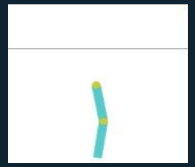
Torque 0

Est.
Torque 1: 100
Torque 0: 45
Torque -1:60

* Est. is the estimate number of frames from the goal

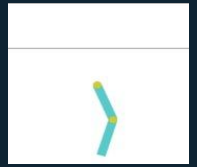
Planning Not Helping?

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Torque 1



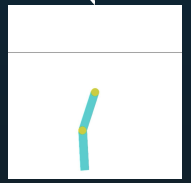
Est.
Torque 1: 80
Torque 0:90
Torque -1:100

Torque 1

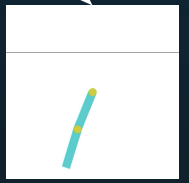
Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

Torque 0

Est.
Torque 1: 100
Torque 0: 45
Torque -1:60



Est.
Torque 1: 80
Torque 0: 70
Torque -1:90

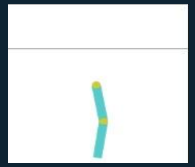


Est.
Torque 1: 90
Torque 0: 70
Torque -1:80

* Est. is the estimate number of frames from the goal

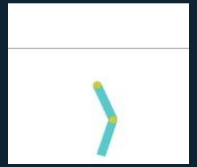
Planning Not Helping?

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Torque 1



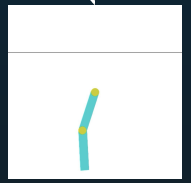
Est.
Torque 1: 80
Torque 0:90
Torque -1:100

Torque 1

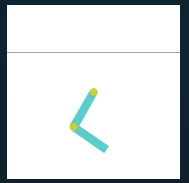
Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

Torque 0

Est.
Torque 1: 100
Torque 0: 45
Torque -1:60



Torque 0



Est.
Torque 1: 80
Torque 0: 70
Torque -1:90

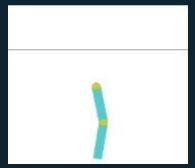
Est.
Torque 1: 70
Torque 0: 60
Torque -1:80

Est.
Torque 1: 90
Torque 0: 70
Torque -1:80

* Est. is the estimate number of frames from the goal

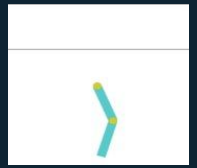
Selective Planning

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Torque 1



Est.
Torque 1: 80
Torque 0:90
Torque -1:100

Torque 1

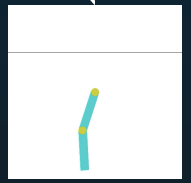
Confident

Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

Torque 0

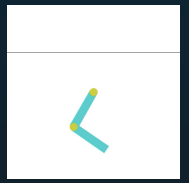
Not Confident

Est.
Torque 1: 100
Torque 0: 45
Torque -1:60

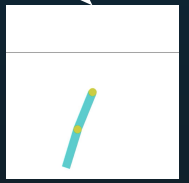


Est.
Torque 1: 80
Torque 0: 70
Torque -1:90

Torque 0



Est.
Torque 1: 70
Torque 0: 60
Torque -1:80

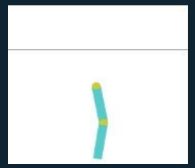


Est.
Torque 1: 90
Torque 0: 70
Torque -1:80

* Est. is the estimate number of frames from the goal

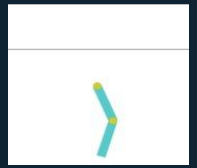
Selective Planning

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Torque 1



Est.
Torque 1: 80
Torque 0:90
Torque -1:100

Torque 1

Confident

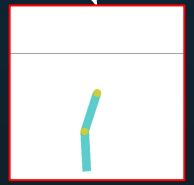
Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

Torque 0

Not Confident

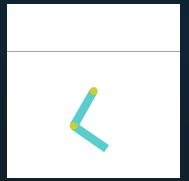
Est.
Torque 1: 100
Torque 0: 45
Torque -1:60

How to measure the uncertainty?

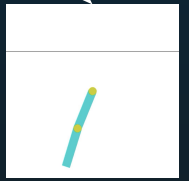


Est.
Torque 1: 80
Torque 0: 70
Torque -1:90

Torque 0



Est.
Torque 1: 70
Torque 0: 60
Torque -1:80

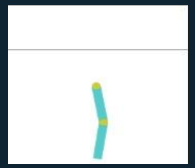


Est.
Torque 1: 90
Torque 0: 70
Torque -1:80

* Est. is the estimate number of frames from the goal

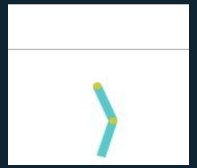
Selective Planning

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Torque 1



Est.
Torque 1: 80
Torque 0:90
Torque -1:100

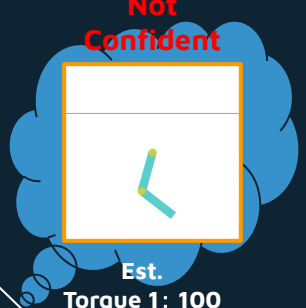
Torque 1

Confident

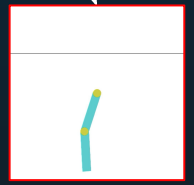
Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

Torque 0

Not Confident

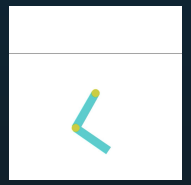


Est.
Torque 1: 100
Torque 0: 45
Torque -1:60

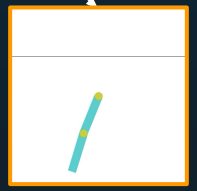


Est.
Torque 1: 80
Torque 0: 70
Torque -1:90

Torque 0



Est.
Torque 1: 70
Torque 0: 60
Torque -1:80



Est.
Torque 1: 90
Torque 0: 70
Torque -1:80

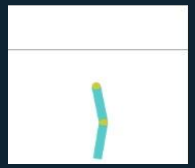
How to measure the uncertainty?

➔ One-Step State Error

* Est. is the estimate number of frames from the goal

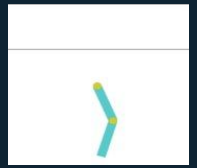
Selective Planning

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

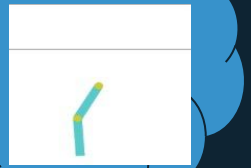
Torque 1



Est.
Torque 1: 80
Torque 0:90
Torque -1:100

Torque 1

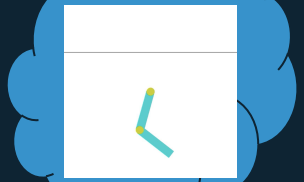
Confident



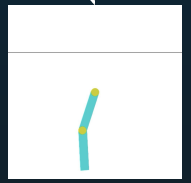
Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

Torque 0

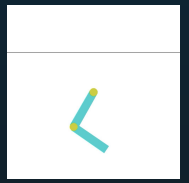
Not Confident



Est.
Torque 1: 100
Torque 0: 45
Torque -1:60



Torque 0



Est.
Torque 1: 80
Torque 0: 70
Torque -1:90

Est.
Torque 1: 70
Torque 0: 60
Torque -1:80

Est.
Torque 1: 90
Torque 0: 70
Torque -1:80

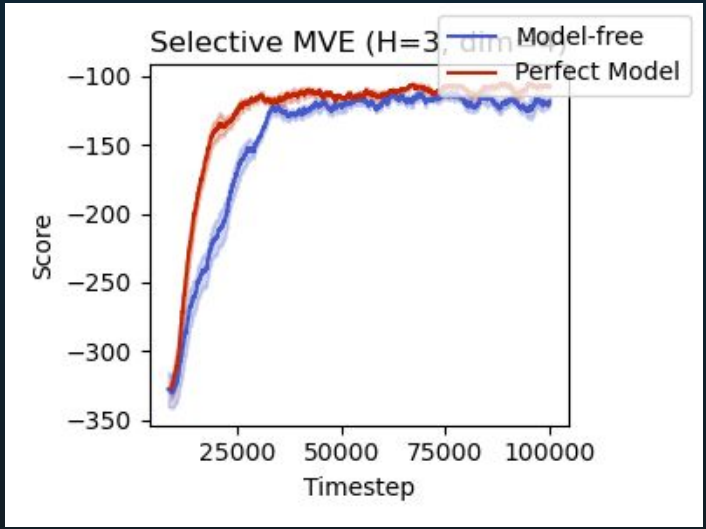
How to measure the uncertainty?

- ➔ One-Step State Error
- ➔ Target Error

* Est. is the estimate number of frames from the goal

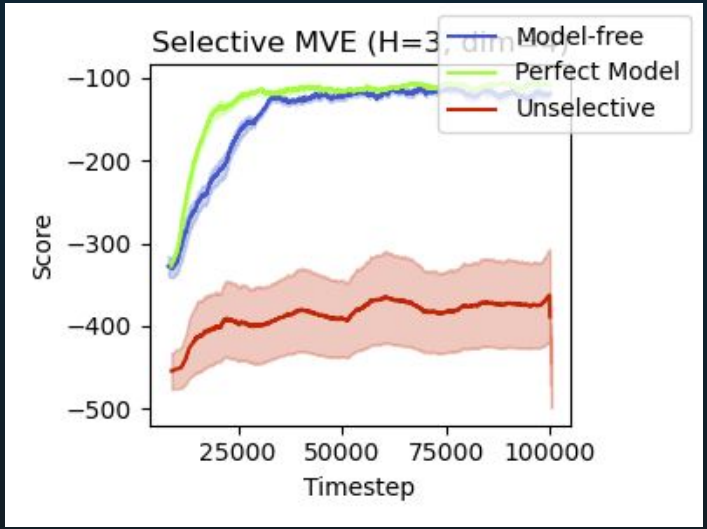
Testing Hypothesis

- State and target can be used as uncertainty signal to measure how confident model is
- Oracle experiment
  - Assume differences are known and weigh experiences differently accordingly
  - For one-step state error, we accumulate them through planning



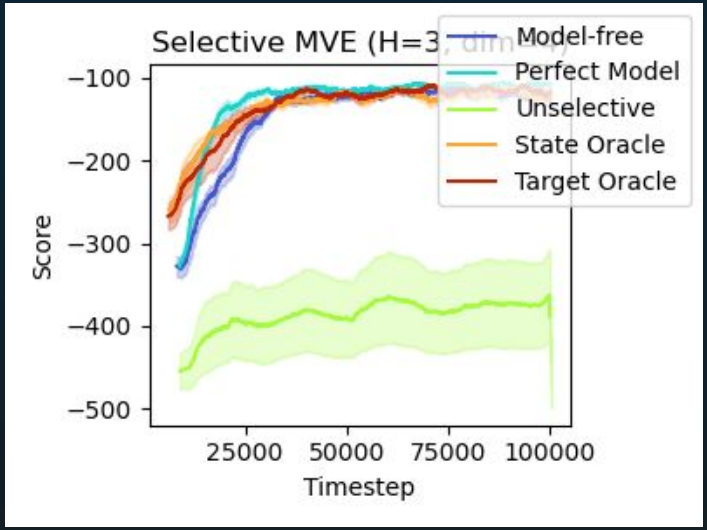
Testing Hypothesis

- State and target can be used as uncertainty signal to measure how confident model is
- Oracle experiment
  - Assume differences are known and weigh experiences differently accordingly
  - For one-step state error, we accumulate them through planning



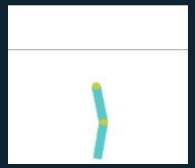
Testing Hypothesis

- State and target can be used as uncertainty signal to measure how confident model is
- Oracle experiment
  - Assume differences are known and weigh experiences differently accordingly
  - For one-step state error, we accumulate them through planning



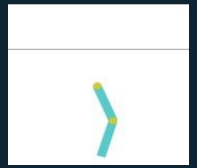
New Idea: Planning With Range

Sample From Real Environment



Est.
Torque 1: 90
Torque 0: 60
Torque -1:90

Torque 1



Est.
Torque 1: 80
Torque 0:90
Torque -1:100

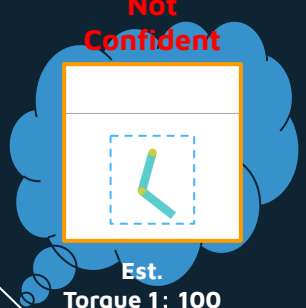
Torque 1

Confident

Est.
Torque 1: 85
Torque 0: 60
Torque -1:95

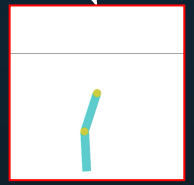
Torque 0

Not Confident



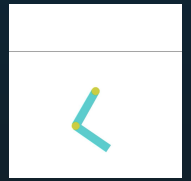
Est.
Torque 1: 100
Torque 0: 45
Torque -1:60

Use range to let the model take uncertainty from last frame into account

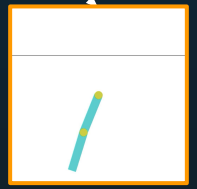


Est.
Torque 1: 80
Torque 0: 70
Torque -1:90

Torque 0



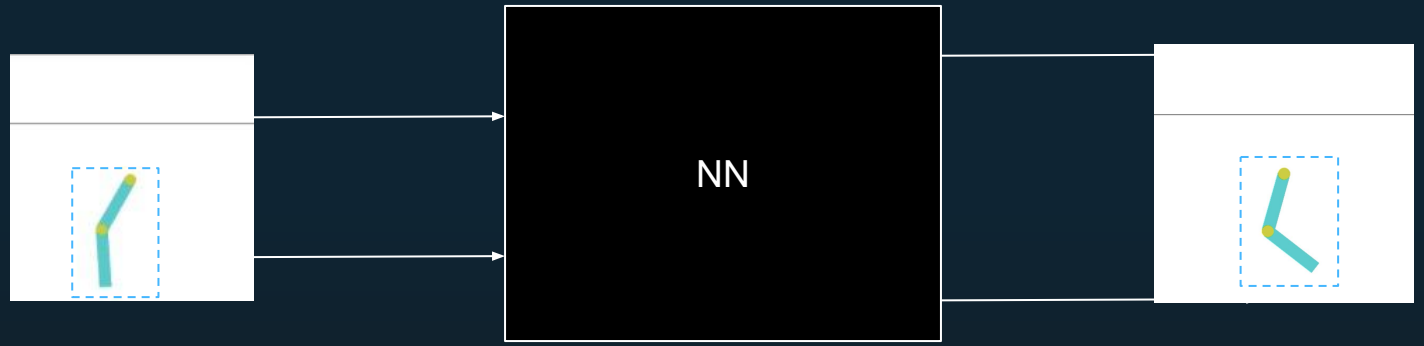
Est.
Torque 1: 70
Torque 0: 60
Torque -1:80



Est.
Torque 1: 90
Torque 0: 70
Torque -1:80

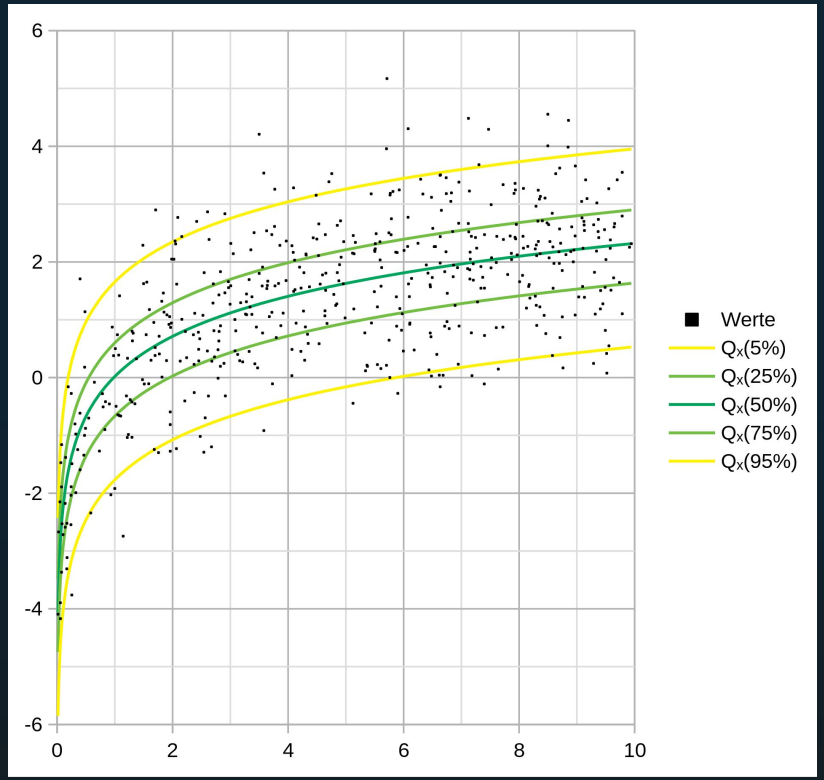
* Est. is the estimate number of frames from the goal

Approach: Neural Networks with Quantile Regression



Approach: Neural Networks with Quantile Regression

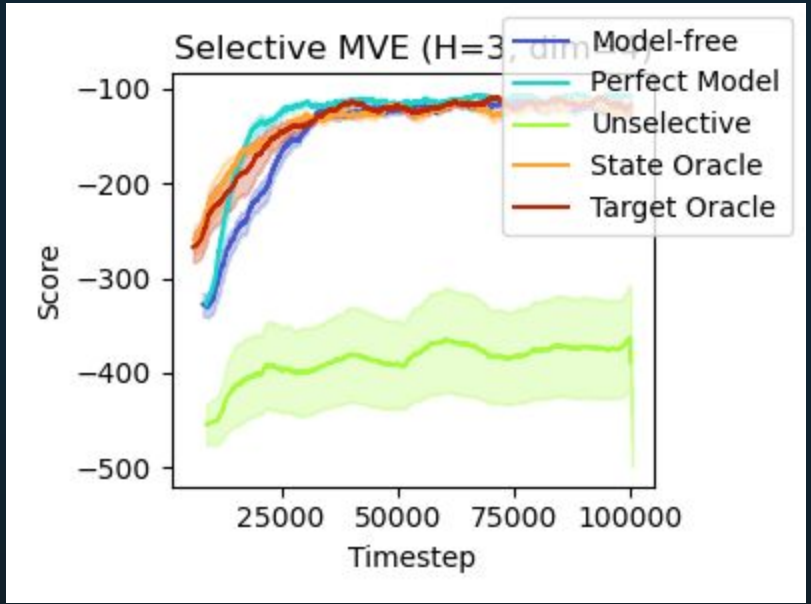
Source: Ready to Score 1500+ on S.A.T? - Boost Your S.A.T Score Quickly



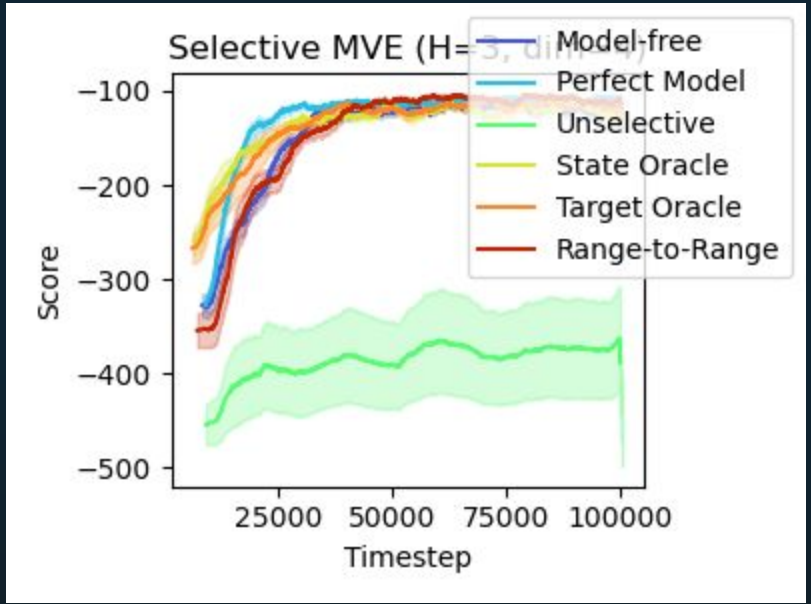
Source: https://en.wikipedia.org/wiki/Quantile_regression#Intuition

Approach: Neural Networks with Quantile Regression

Preliminary Results



Preliminary Results



Reference

- MBRL Wishlist, Talk by Erin Talvitie
- Sutton, Richard S., and Andrew G. Barto. Reinforcement Learning: An Introduction. The MIT Press, 2012.
- Abbas, Zaheer, et al. “Selective Dyna-Style Planning under Limited Model Capacity.” arXiv.Org, 7 Mar. 2021, arxiv.org/abs/2007.02418.