Assignment 2: Policy Gradient

Andrew ID: `kkabeer`
Collaborators: `asenathi`
NOTE: Please do **NOT** change the sizes of the answer blocks or plots.

5 Small-Scale Experiments

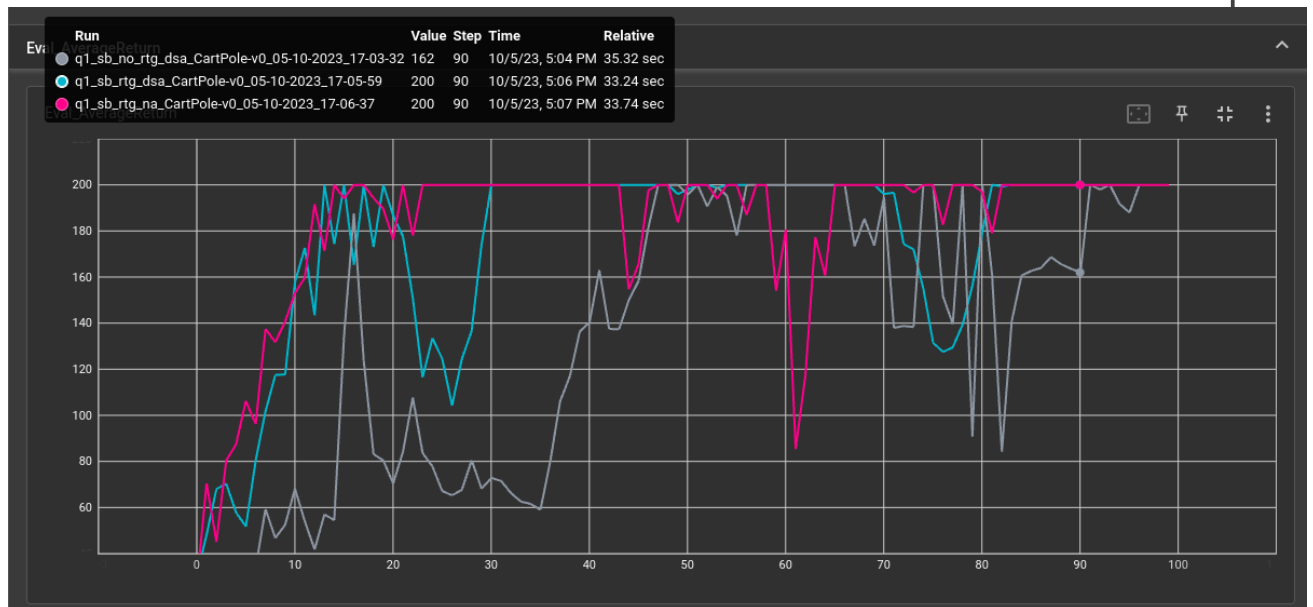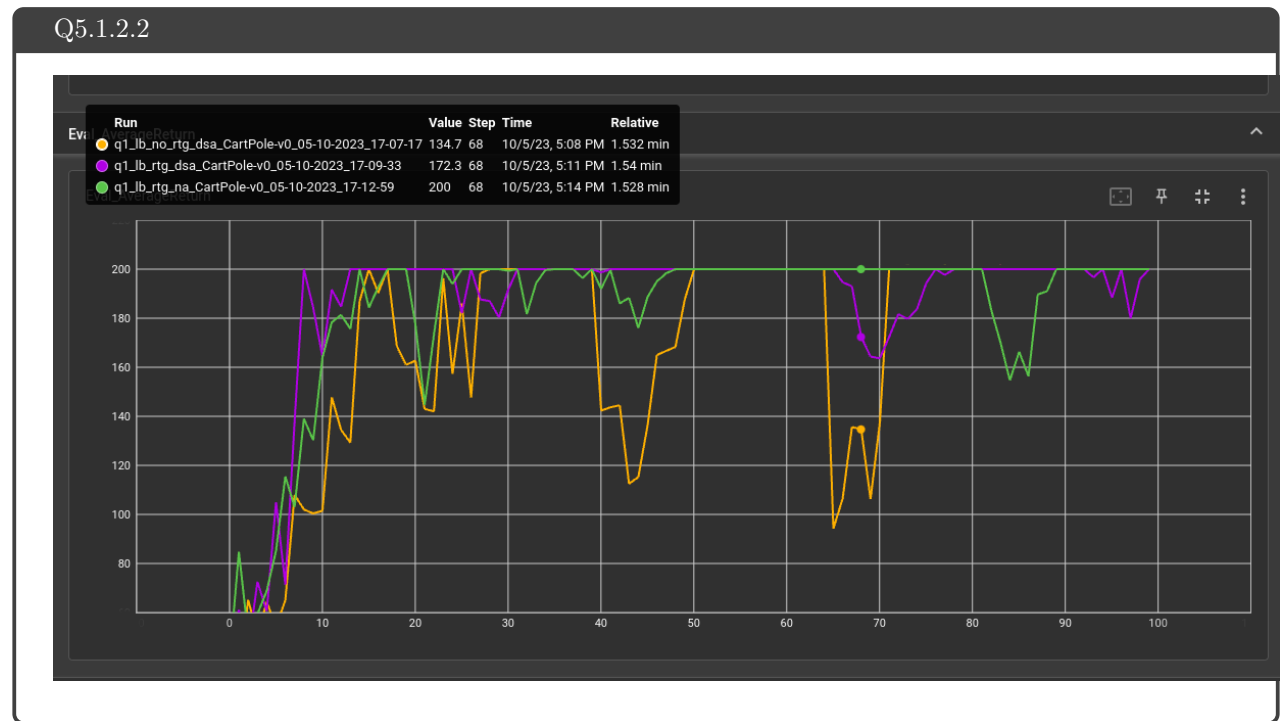## 5.1 Experiment 1 (Cartpole) – [25 points total]

5.1.1 Configurations

> Q5.1.1

5.1.2 Plots

5.1.2.1 Small batch – [5 points]

> Q5.1.2.1
>
>

5.1.2.2 Large batch – [5 points]

Q5.1.2.2

5.1.3 Analysis

5.1.3.1 Value estimator – [5 points]

Q5.1.3.1

5.1.3.2 Advantage standardization – [5 points]

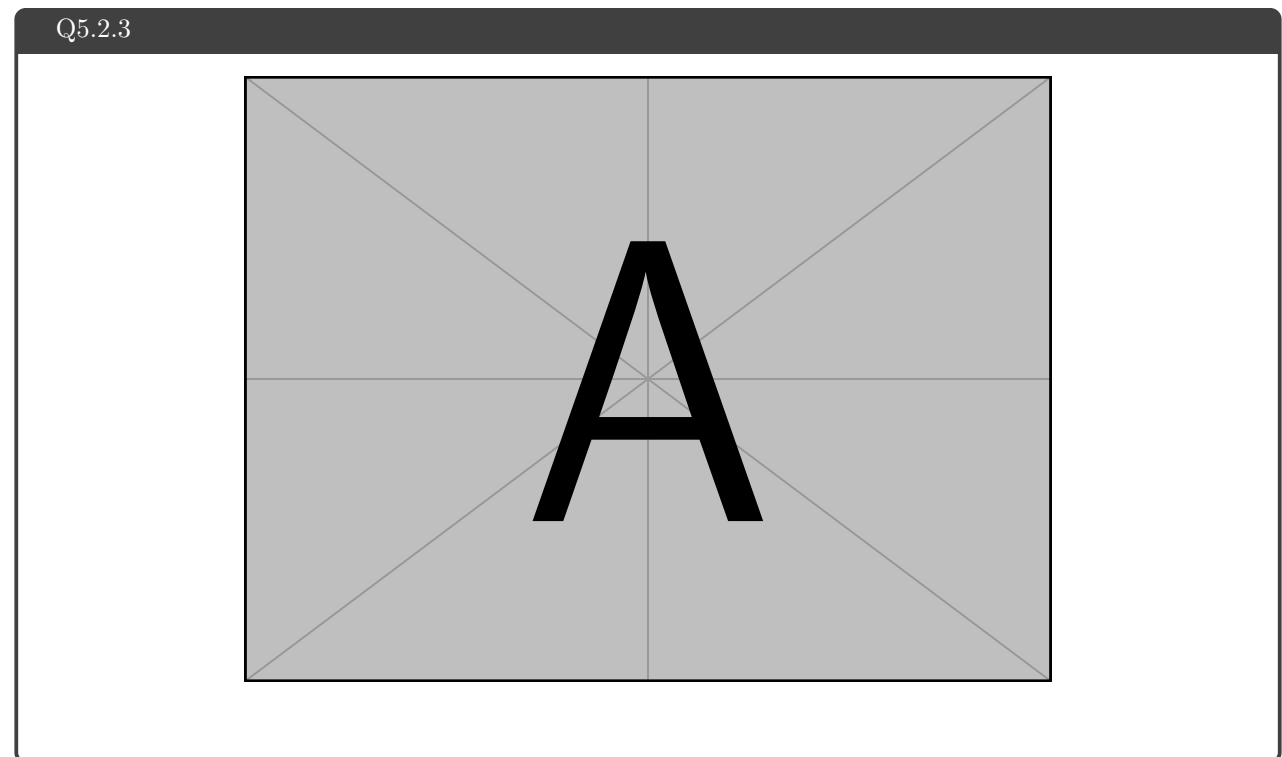**Q5.1.3.2**

5.1.3.3 Batch size – [5 points]
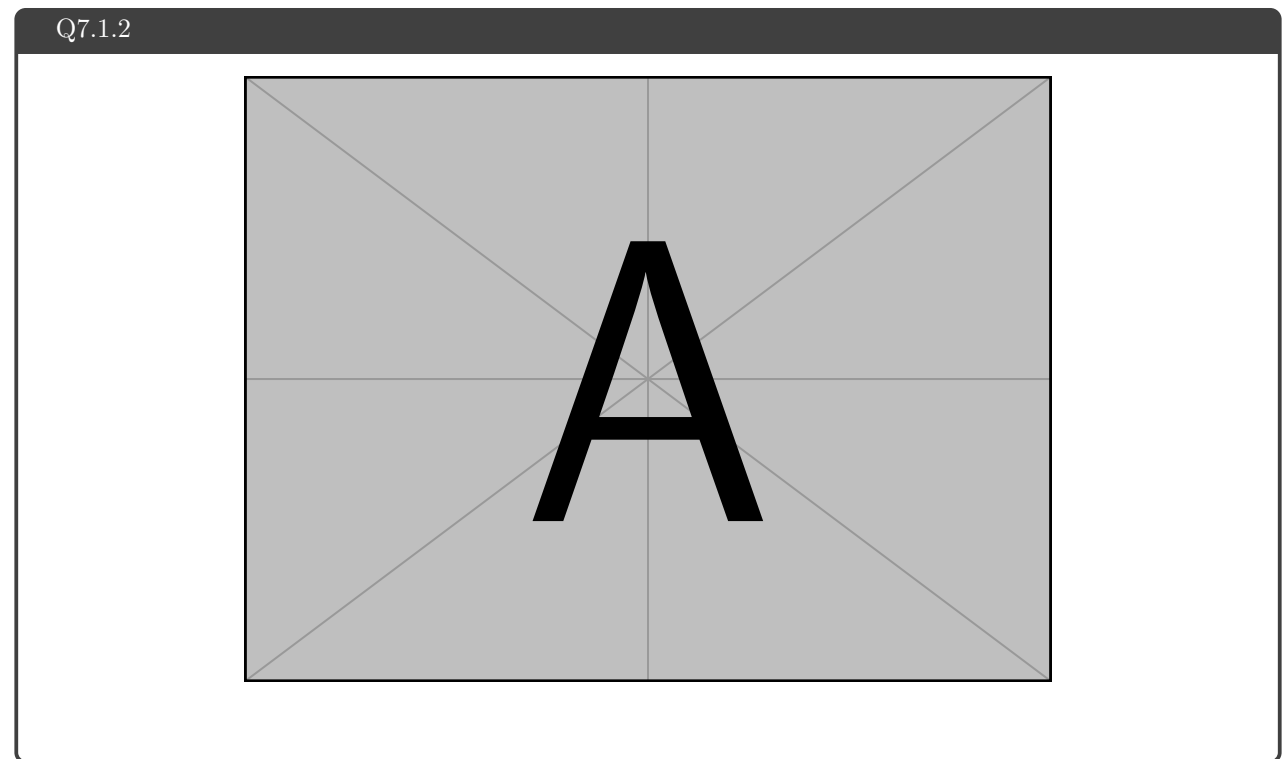
> **Q5.1.3.1**
>
>

5.2 Experiment 2 (InvertedPendulum) – [15 points total]

5.2.1 Configurations – [5 points]

> **Q5.2.1**
>
>

5.2.2 smallest b* and largest r* (same run) – [5 points]

> **Q5.2.2**
>
>

5.2.3 Plot – [5 points]

Q5.2.3

7 More Complex Experiments

7.1 Experiment 3 (LunarLander) – [10 points total]

7.1.1 Configurations

Q7.1.1

7.1.2   Plot – [10 points]

Q7.1.2

7.2   Experiment 4 (HalfCheetah) – [30 points]

7.2.1   Configurations

Q7.2.1

7.2.2   Plot – [10 points]

> **Q7.2.2**
>
>
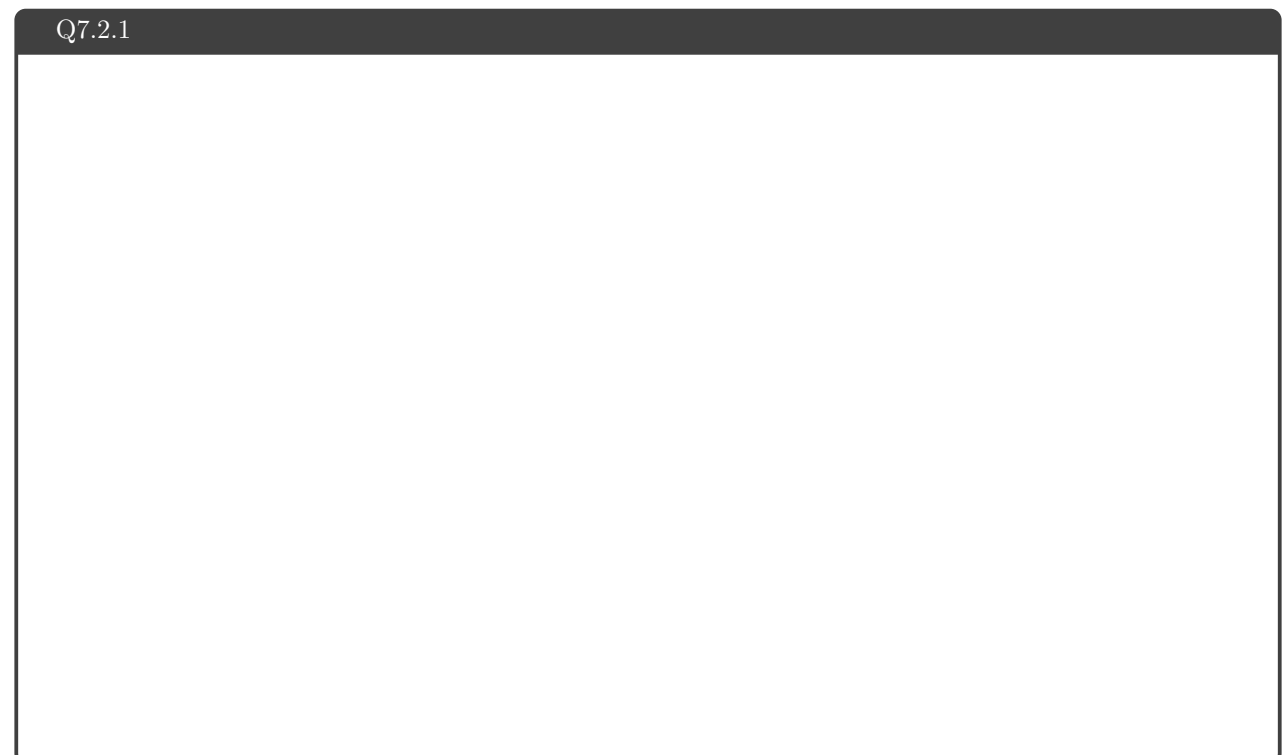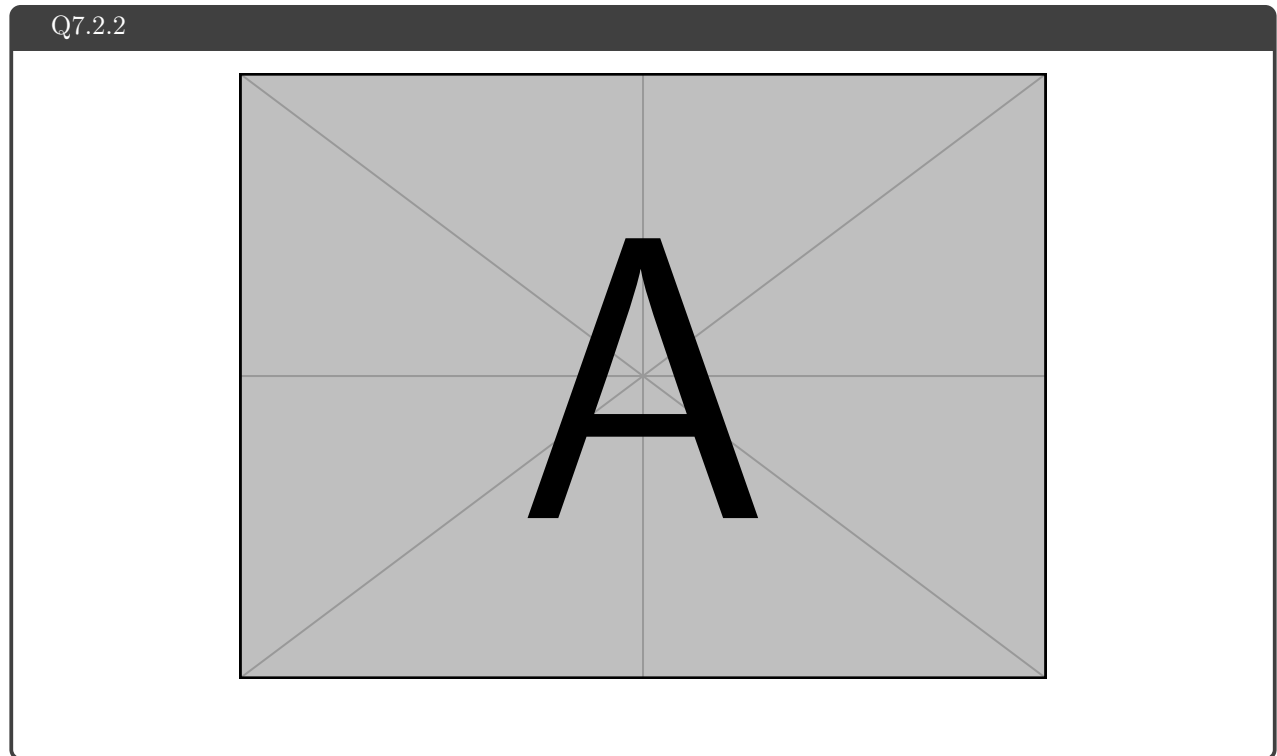
7.2.3   Optimal b* and r* – [3 points]

> **Q7.2.3**

7.2.4   Describe how b* and r* affect task performance – [7 points]

> **Q7.2.4**

7.2.5   Configurations with optimal b* and r* – [3 points]

> **Q7.2.5**
>
>

7.2.6   Plot for four runs with optimal b* and r* – [7 points]

> **Q7.2.6**
>
>
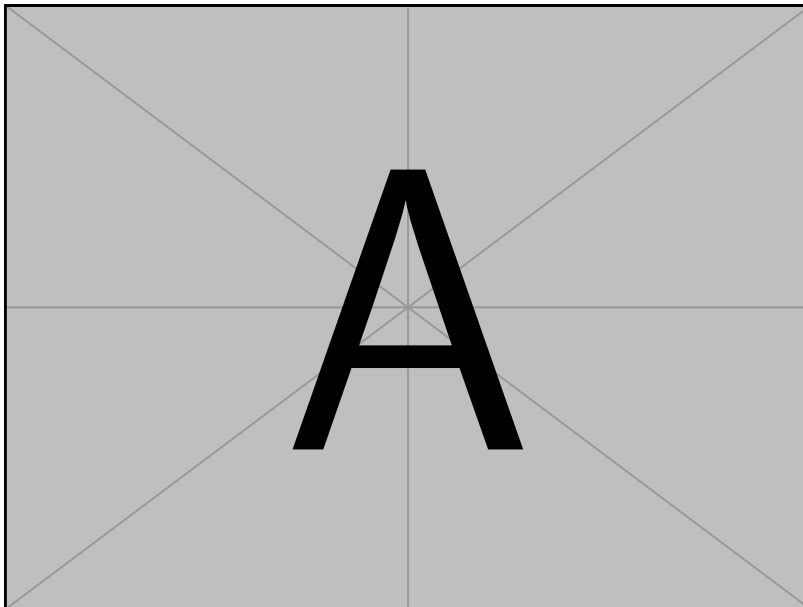
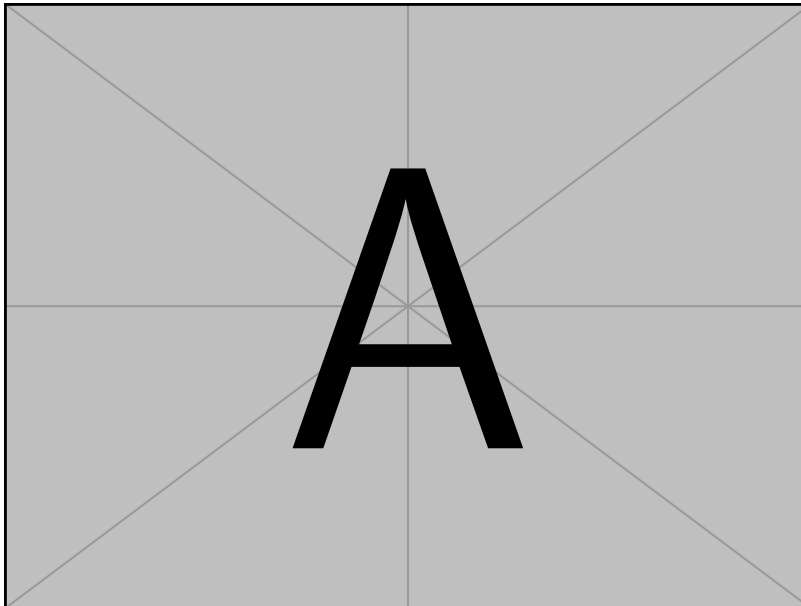## 8   Implementing Generalized Advantage Estimation

8.1 Experiment 5 (Hopper) – [20 points]

8.1.1 Configurations

> **Q8.1.1**

8.1.2 Plot – [13 points]

> **Q8.1.2**
>
>

8.1.3 Describe how λ affects task performance – [7 points]

> **Q8.1.3**

9   Bonus! (optional)

9.1   Parallelization – [15 points]

> **Q9.1**
>
> Difference in training time:

9.2   Multiple gradient steps – [5 points]

> **Q9.1**
>
>