

CS 4320: Machine Learning

Assignment: Temporal Difference Q-Learning (Reinforcement Learning)

Train a reinforcement agent to perform in the `Taxi-v3` environment.

It is expected that you will use the `FrozenLake` example code as a starting point for your code development.

Search the hyper parameter space to identify the best combination of values for alpha, gamma, and epsilon-chance-factor. Best is determined by the average reward during training epochs, and the average reward during scoring epochs. Note, this means you'll be reporting 2 combinations.

Use 5000 learning epochs and 100 scoring epochs.

Create a report that includes:

- A description of the states, actions, and rewards of the `Taxi-v3` environment.
- An outline of code changes needed to work with the `Taxi-v3` environment.
- A description of the search process used.
- A table of representative samples from the hyper parameter search space, with the learning reward and scoring rewards. You may do hundreds or even thousands of samples from the space. Only include 10-20 that represent the results. Mark the samples that obtain the best learning and scoring rewards.
- A discussion of the effect of hyper parameter values on the average rewards.

Required Steps

- Download the starter code.
- Verify it runs for the FrozenLake example.
- Modify the code to run with the Taxi example.
- Explore the hyper parameter space, keeping results for use in the report.
- Create a report with the contents mentioned above.
- Commit and push your code in the git repository.
- Submit the report (as PDF) to Canvas.