Source Code:

mypa3_mf_control.py:

I have implemented MC Control, SARSA, & Q-Learning as per the pseudocode and the code template provided.

- About MC: I have experimented with generating episodes that terminate at the terminal state, but this approach has proven to be incredible time-consuming for meager differences in best runs. I have opted to terminate episodes at 500 steps and perform updates to the Q-table on incomplete episodes.
- About SARSA: Aside from minor changes, its the same as my implementation from CS156
- About Q-Learning: Direct translation of the provided pseudocode.

mymain_sol.py:

A copy, paste, tweak of the provided sample code to run our agents. I have modified mine so that it iterates over the different environments I have selected. For each environment an agent runs (train from scratch) and provides its best return 5 times. I have recorded all of my output.

Output Logs:

Honestly, feel free to ignore the logs and go straight to the table of results I have organized below.

5k_eps_out.txt: Output for running each agent on each environment 5 times with 5 thousand training episodes.

15k_eps_out.txt: Output for running each agent on each environment 5 times with 15 thousand training episodes.

Results (next page):

MC is not so great because generating episodes is cumbersome.

SARSA and Q-Learning seem to struggle in the stochastic environment presented in frozen lake, as I have observed that turning off ‘slippery’ tends to improve returns.

Further improvements could be attained by tweaking learning parameters (step size, total eps, or decaying epsilon).

CliffWalking-v0						
	MC (5k eps)	MC (15k eps)	SARSA (5k eps)	SARSA (15k eps)	Q-Learning (5k eps)	Q-Learning (15k eps)
Run #1	-14.53391169	-14.53391169	-13.07154487	-14.53391169	-11.54888054	-11.54888054
Run #2	-15.93836879	-14.53391169	-13.07154487	-13.07154487	-11.54888054	-11.54888054
Run #3	-14.53391169	-14.53391169	-13.07154487	-14.53391169	-11.54888054	-11.54888054
Run #4	-14.53391169	-14.53391169	-13.07154487	-14.53391169	-11.54888054	-11.54888054
Run #5	-14.53391169	-14.53391169	-14.53391169	-14.53391169	-11.54888054	-11.54888054
Average Return/Run	-14.81480311	-14.53391169	-13.36401823	-14.24143833	-11.54888054	-11.54888054

FrozenLake-v1						
	MC	MC (15k eps)	SARSA	SARSA (15k eps)	Q-Learning	Q-Learning (15k eps)
Run #1	0	0	0	0.2110609207	0	0.3291805474
Run #2	0	0	0	0	0	0.7690223893
Run #3	0	0	0.7237977206	0.567976176	0	0.1409059532
Run #4	0.5795675265	0	0	0.1946758731	0	0.3716017144
Run #5	0.6542558123	0.3497485608	0	0	0	0
Average Return/Run	0.2467646678	0.06994971215	0.1447595441	0.1947425939	0	0.3221421208

Taxi-v3						
	MC	MC (15k eps)	SARSA	SARSA (15k eps)	Q-Learning	Q-Learning (15k eps)
Run #1	6.051194552	-43.36902221	4.930170661	1.699837185	2.754935903	0.6658404415
Run #2	10.76878733	7.195096482	8.362343349	2.754935903	7.195096482	8.362343349
Run #3	-43.36902221	6.051194552	-43.36902221	0.6658404415	4.930170661	1.699837185
Run #4	1.699837185	-2.313716303	-43.36902221	3.831567248	6.051194552	-0.3474763674
Run #5	8.362343349	-2.313716303	-43.36902221	-43.36902221	7.195096482	10.76878733
Average Return/Run	-3.297371958	-6.950032755	-23.36291052	-6.883368285	5.625298816	4.229866387