Optional Extra Assignment

This optional assignment question A5.1 allows you to earn 8 extra points, so that you can still take the exam if it brings your final sum of all points to above 75%, i.e., to at least 51 points. The solution must be submitted on Brightspace before the deadline (see due date above). Please submit all answers in one PDF file.

You will be asked to implement and test something in python/pytorch. We recommend that you use a Jupyter Notebook, convert your final version (including result plots) to PDF ("Download as" → "PDF via LaTeX") and attach the PDF at the end of your submission PDF. You are welcome to use other editors, but please make sure you submit the code (or the crucial code segments) and the results in an easily readable PDF format. Unreadable files will yield no points.

Good luck!

A5.1: Policy gradients and exploration **(8 points)**

Run the PPO implementation from A3.1 in assignment sheet 3 (yours or the sample solution's) on the environments `Mountaincar-v0` and `Acrobot-v1` for 500k steps, both with a maximal episode length of 200. As in A4.1 of assignment sheet 4, the algorithm should only rarely be able to learn anything. To allow learning, extend the PPO algorithm with intrinsic exploration reward based on RND novelty/uncertainty estimation as in exercise sheet 4.

Only if your code implements intrinsic reward correctly into PPO, and your plots show some signs of learning, you will receive the 8 extra points.

Total 8 points.