Homework 5

Nicklas Hansen (s153077) s153077@student.dtu.dk
Peter Ebert Christensen (s153758) pebch@dtu.dk

June 19, 2019

Attribution table

Part	Responsible
Implementation	s153077 (49%), s153758 (51%)
Figure 1 & 2	s153077
Figure 3 & 4	s153758

Problem 1

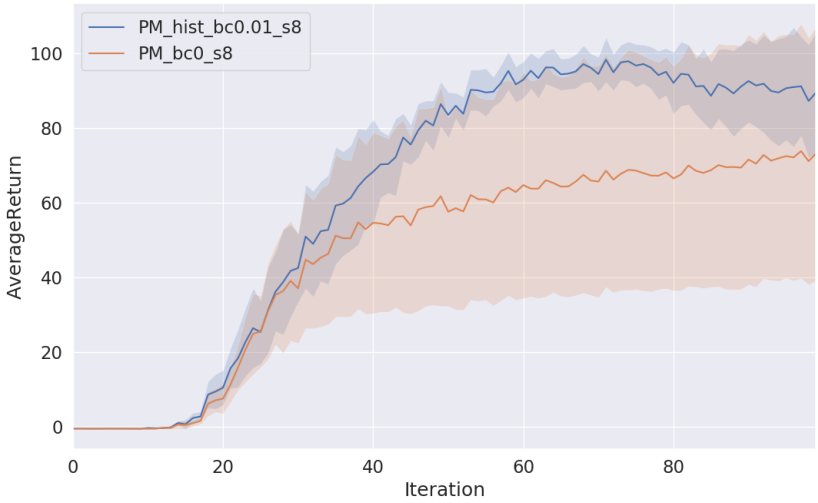

Figure 1: *Model predictions for PointMass environment:* Plot of average return for the histogram density model (blue) and the count-based reward bonus model without exploration (red). It is observed that a histogram-based exploration achieves a higher average return and has far smaller variance during training.

Problem 2

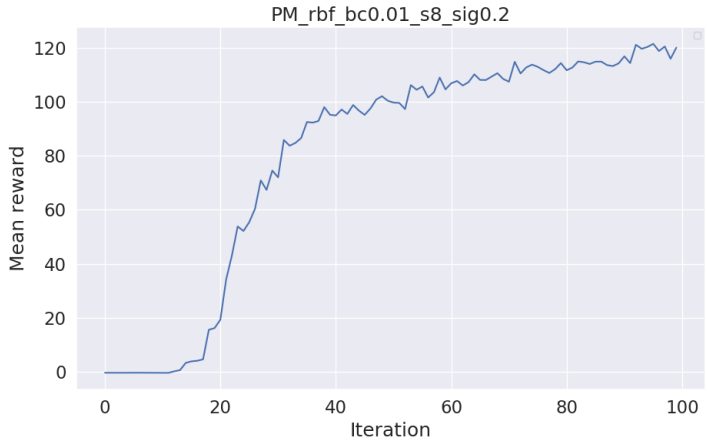

Figure 2: *Model predictions for PointMass environment:* Plot of average return for using the RBF density model with parameters as defined in the problem description. Due to issues with the plot function for this particular run, we re-implemented a simple plot for illustrative purposes. It can be concluded that RBF works really well for exploration (obtained the greatest average return out of all density models considered), but its memory requirements (roughly 12 GB) and wall-times are prohibitively large for use in more complex environments.

Problem 3

Figure 3: *Model predictions for PointMass environment:* Plot of returns of the EX2 discriminator: KDE based exploration (blue) and the count based reward bonus model without exploration (red). It can be concluded that the EX2 discriminator is quite stable as it has a consistent low variance in every iteration compared to the model without exploration. However it remains unclear if it has any significant performance advantage over the histogram model.

Problem 4

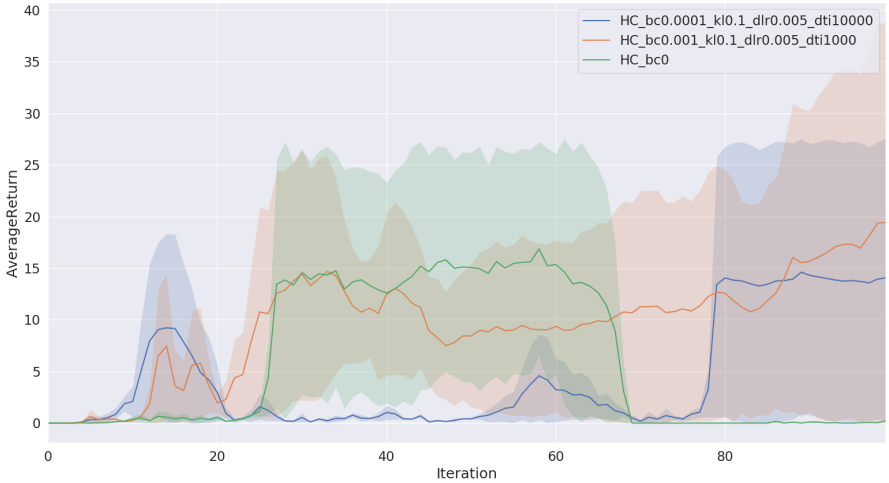

Figure 4: *Model predictions for HalfCheetah environment* Plot of returns of the EX2 discriminator with different bonus coefficients (0.0001 and 0.001) and number of training episodes (10000 and 1000)(blue, red) and the count based reward bonus model without exploration (green). We can conclude that the a low bonus coefficient the model will be as good as model without exploration and with the right parameters it will be consistent in its exploration and achieve larger returns over time.