

Ege Can KAYA
2018400018

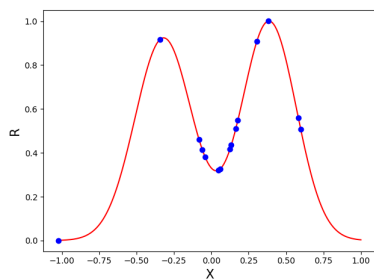
CmpE 343: Introduction to Probability and Statistics
for Computer Engineers (Fall 2020)
Bonus Homework

In the solution to the presented reinforcement learning problem, I made use of a gradient-based approach and the Gaussian distribution. With each reward function, I iterated the sampling process 5 times, and took a sample of 15 points each time.

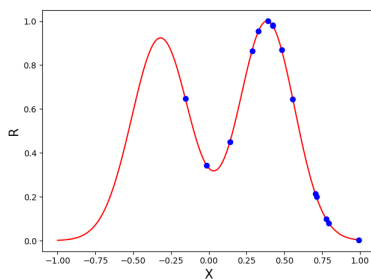
In the first sample, I took 0 as the mean of my Gaussian distribution to start out with a central value, and took a rather large standard deviation of 0.4, in order to make sure that the sampled points covered a large interval. It is logical to take a sample from a large interval initially, in order to get a good sense of where the reward function assumes its highest values. This also handles the problem of getting stuck at a local maximum, as opposed to the global maximum.

In the iterations following the first one, my method was as follows: I determined the point where the reward function had the highest value among our sampled points, and chose that point as the new center of our Gaussian sample. I also decreased the standard deviation gradually with each iteration, to restrict our sample to a small interval around the highest value. To be exact, I decreased the value of the standard deviation by 0.13 at each of the first three iterations, and set it equal to 0 in the final two iterations. This method was very successful in finding the global maximum of the reward function, namely, it succeeded in correctly identifying the global maximum in 94 of the 100 random trials I have done. Only in six trials, where the value of the global maximum and the local maximum were very close, did the samples get stuck around the local maximum, failing to identify the global maximum.

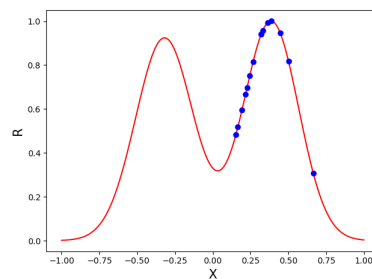
In the following pages are figures for the sampled points in five successful iterations, for five different environments.



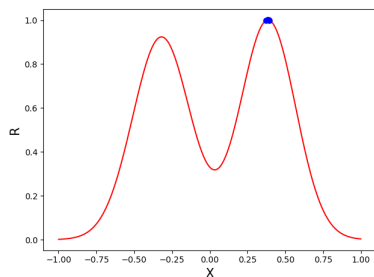
(a) Iteration 1



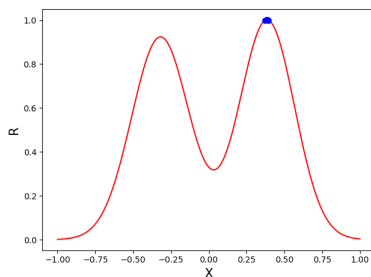
(b) Iteration 2



(c) Iteration 3

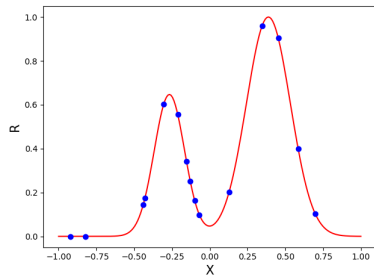


(d) Iteration 4

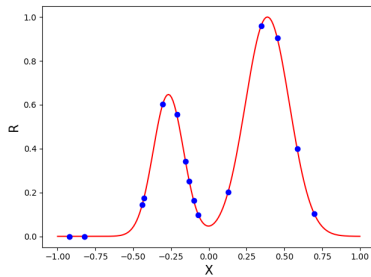


(e) Iteration 5

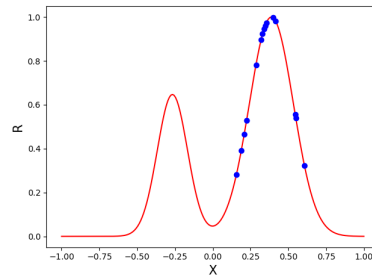
Figure 1: Environment 1



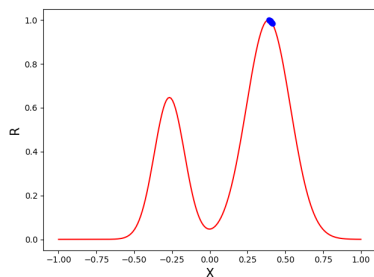
(a) Iteration 1



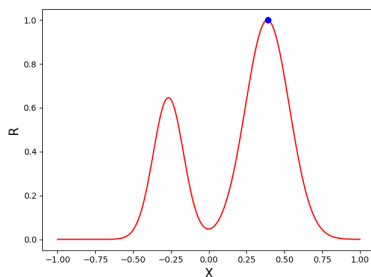
(b) Iteration 2



(c) Iteration 3

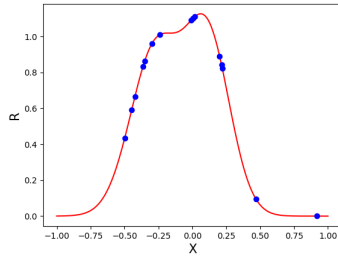


(d) Iteration 4

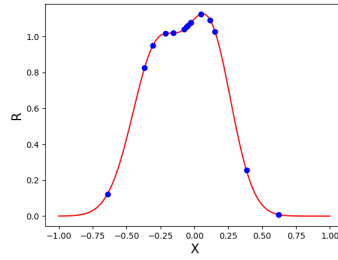


(e) Iteration 5

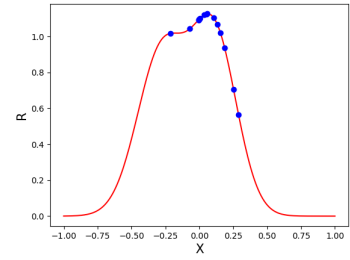
Figure 2: Environment 2



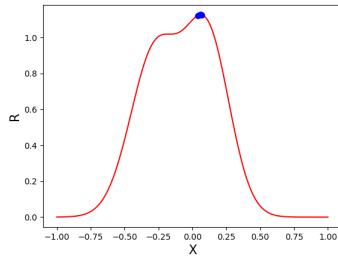
(a) Iteration 1



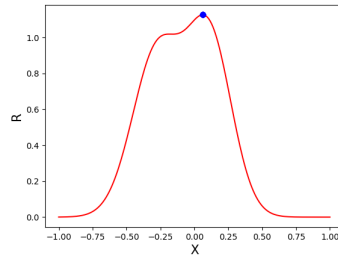
(b) Iteration 2



(c) Iteration 3

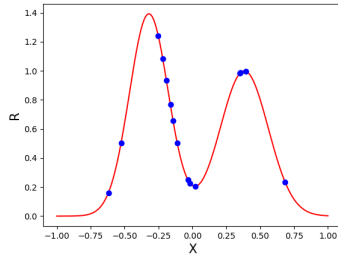


(d) Iteration 4

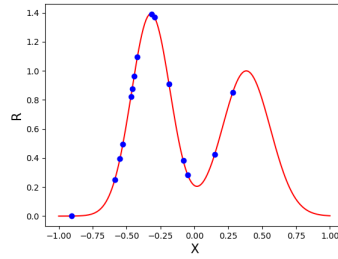


(e) Iteration 5

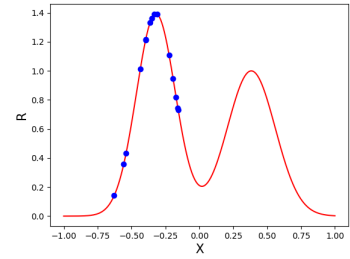
Figure 3: Environment 3



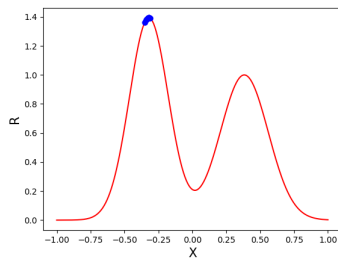
(a) Iteration 1



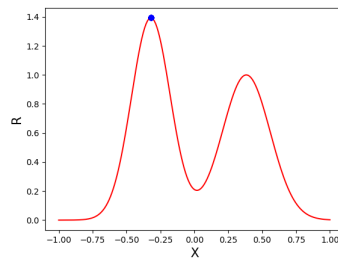
(b) Iteration 2



(c) Iteration 3

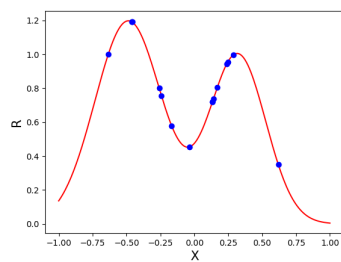


(d) Iteration 4

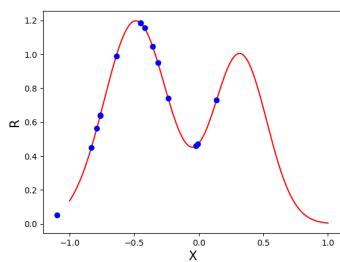


(e) Iteration 5

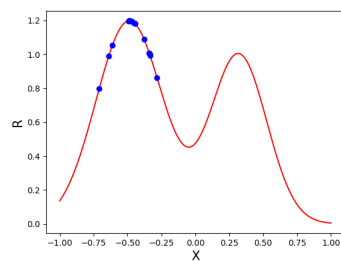
Figure 4: Environment 4



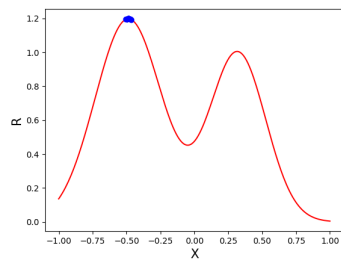
(a) Iteration 1



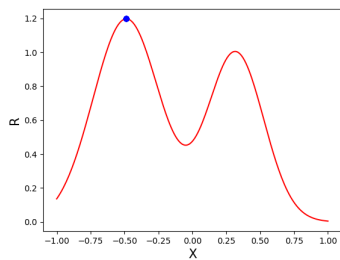
(b) Iteration 2



(c) Iteration 3



(d) Iteration 4



(e) Iteration 5

Figure 5: Environment 5