Homework Assignment 7

Zhaoyang Xu

November 1, 2020

Contents

1 CMA-ES for Reinforcement Learning (50 points)

1.1

```
Iterat #Fevals   function value  axis ratio  sigma  min&max std  t[m:s]
    1     15 -2.000000000000000e+02 1.0e+00 9.50e-03  9e-03  1e-02 0:00.2
    2     30 -1.090000000000000e+02 1.0e+00 9.10e-03  9e-03  9e-03 0:00.3
    3     45 -1.590000000000000e+02 1.0e+00 8.80e-03  9e-03  9e-03 0:00.4
    5     75 -1.000000000000000e+03 1.1e+00 8.57e-03  9e-03  9e-03 0:00.9
termination on ftarget=-999.9 (Sun Nov  1 14:31:29 2020)
final/bestever f-value = -8.900000e+01 -1.000000e+03
incumbent solution: [-0.01608988 -0.01919164 -0.00680728  0.01004572
0.02937099 -0.0082997
  0.01206515 -0.00917344 ...]
std deviations: [0.00854533 0.00854307 0.00853354 0.00855662 0.00857953
0.00856339
 0.00853891 0.00851933 ...]
```

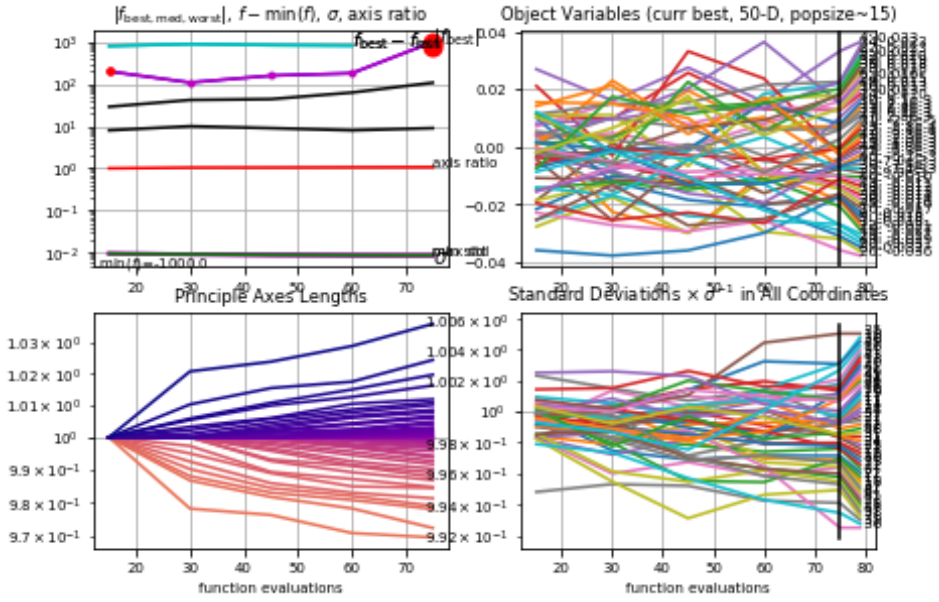

Figure 1: **Figure of "Neuroevolution Strategies" approach**

1.2

For RNN, we have an initial parameter a_1. By given weight for a_1 and x_1 separately, we can have $w_a a_1$ and $w_x x_1$. Then by function tanh, we can have the next parameter a_2.

```
p_1 = np.matmul(w_a,a_1)
p_2 = np.matmul(w_x,x_1)
a_2 = np.tanh(p_1+p_2)
```

Then, repeat the steps, we can get a_3 as the action.

1.3

If have bias, the learning will become more difficult. Like for 1.1, if we add bias, it will iterate more times to get the answer.

2 UCLR2 Revisited (30 points)

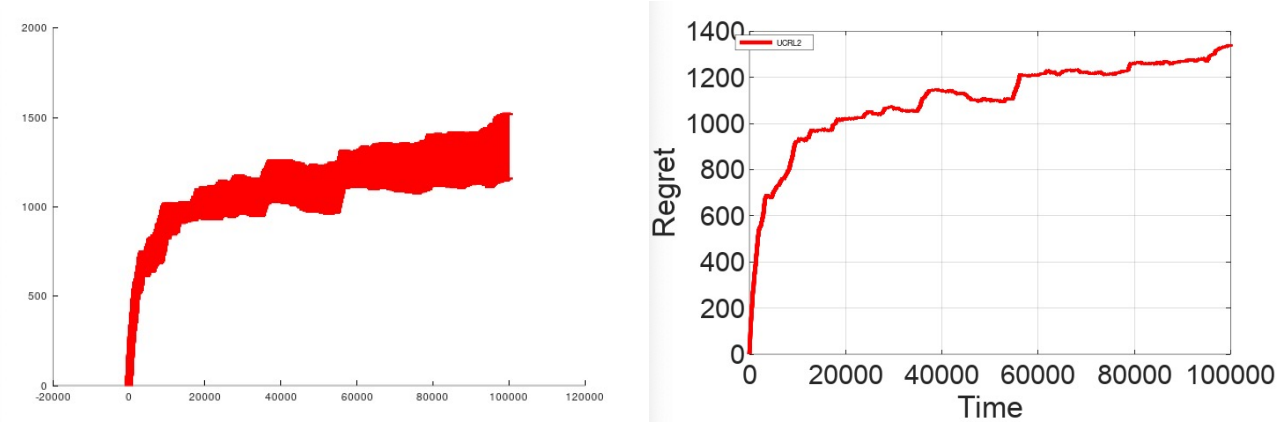

Figure 2: **UCRL2-L of 6-states**

3 Computing Diameter (20 points)

$$D := \max_{s \neq s'} \min_{\pi} \mathbb{E}[T^{\pi}(s', s)] = \sum \frac{1}{\delta}$$

$L = 6$	$D = 14.72$
$L = 12$	$D = 34.72$
$L = 25$	$D = 78.06$