

```
1 "/Users/rohanpatel/Desktop/github_repo/
  Reinforcement Learning/env/bin/python" /Users/
  rohanpatel/Desktop/github_repo/Reinforcement
  Learning/Atari DQN Project/Linear.py
2 Starting Training
3 layer weights: Parameter containing:
4 tensor([[ 0.8567],
5          [-0.6501],
6          [ 0.7370]], dtype=torch.float64,
7          requires_grad=True)
7 layer weights: Parameter containing:
8 tensor([[ 0.7999],
9          [-0.4384],
10         [-0.6402]], dtype=torch.float64,
11         requires_grad=True)
11 Summary AFTER training
12
13 =====
14 Summary:
15 State    Action   Next State   Reward
16 0        0        Q(s,a)= tensor(0.1459, dtype=
17 torch.float64, grad_fn=<SelectBackward0>)
17 0        1        Q(s,a)= tensor(0.1174, dtype=
18 torch.float64, grad_fn=<SelectBackward0>)
18 0        2        Q(s,a)= tensor(0.1228, dtype=
19 torch.float64, grad_fn=<SelectBackward0>)
19 1        0        Q(s,a)= tensor(0.1989, dtype=
20 torch.float64, grad_fn=<SelectBackward0>)
20 1        1        Q(s,a)= tensor(0.0953, dtype=
21 torch.float64, grad_fn=<SelectBackward0>)
21 1        2        Q(s,a)= tensor(0.1171, dtype=
22 torch.float64, grad_fn=<SelectBackward0>)
22 2        0        Q(s,a)= tensor(0.2520, dtype=
23 torch.float64, grad_fn=<SelectBackward0>)
23 2        1        Q(s,a)= tensor(0.0732, dtype=
24 torch.float64, grad_fn=<SelectBackward0>)
24 2        2        Q(s,a)= tensor(0.1115, dtype=
25 torch.float64, grad_fn=<SelectBackward0>)
25 3        0        Q(s,a)= tensor(0.3050, dtype=
26 torch.float64, grad_fn=<SelectBackward0>)
26 3        1        Q(s,a)= tensor(0.0511, dtype=
```

```
26 torch.float64, grad_fn=<SelectBackward0>)
27 3          2          Q(s,a)= tensor(0.1059, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
28 4          0          Q(s,a)= tensor(0.3580, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
29 4          1          Q(s,a)= tensor(0.0290, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
30 4          2          Q(s,a)= tensor(0.1002, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
31 =====
32
33 Best trajectory:
34 Best trajectory from Test Environment
35 State: tensor([0.], dtype=torch.float64) Action
 : 0 Reward Received: 0
36 State: tensor([1.], dtype=torch.float64) Action
 : 0 Reward Received: 0
37 State: tensor([2.], dtype=torch.float64) Action
 : 0 Reward Received: 0
38 State: tensor([3.], dtype=torch.float64) Action
 : 0 Reward Received: 1
39
40 Total Reward Received: 1
41 Taking a look at model parameters to see if weights
   are changing
42 Parameter containing:
43 tensor([[ 0.0530],
44           [-0.0221],
45           [-0.0056]], dtype=torch.float64,
   requires_grad=True)
46 Parameter containing:
47 tensor([0.1459, 0.1174, 0.1228], dtype=torch.
   float64, requires_grad=True)
48 Starting Training
49 layer weights: Parameter containing:
50 tensor([[ 0.0620],
51           [-0.8060],
52           [ 0.8874]], dtype=torch.float64,
   requires_grad=True)
53 layer weights: Parameter containing:
54 tensor([[ 0.4737],
```

```
55      [ 0.9847],  
56      [-0.9481]], dtype=torch.float64,  
  requires_grad=True)  
57 Summary AFTER training  
58  
59 =====  
60 Summary:  
61 State Action Next State Reward  
62 0      0      Q(s,a)= tensor(0.1189, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
63 0      1      Q(s,a)= tensor(0.1003, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
64 0      2      Q(s,a)= tensor(0.1022, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
65 1      0      Q(s,a)= tensor(0.1550, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
66 1      1      Q(s,a)= tensor(0.0766, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
67 1      2      Q(s,a)= tensor(0.0964, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
68 2      0      Q(s,a)= tensor(0.1911, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
69 2      1      Q(s,a)= tensor(0.0528, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
70 2      2      Q(s,a)= tensor(0.0906, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
71 3      0      Q(s,a)= tensor(0.2272, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
72 3      1      Q(s,a)= tensor(0.0291, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
73 3      2      Q(s,a)= tensor(0.0848, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
74 4      0      Q(s,a)= tensor(0.2633, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
75 4      1      Q(s,a)= tensor(0.0054, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
76 4      2      Q(s,a)= tensor(0.0790, dtype=  
  torch.float64, grad_fn=<SelectBackward0>)  
77 =====  
78  
79 Best trajectory:
```

```
80 Best trajectory from Test Environment
81 State: tensor([0.], dtype=torch.float64) Action
     : 0 Reward Received: 0
82 State: tensor([1.], dtype=torch.float64) Action
     : 0 Reward Received: 0
83 State: tensor([2.], dtype=torch.float64) Action
     : 0 Reward Received: 0
84 State: tensor([3.], dtype=torch.float64) Action
     : 0 Reward Received: 1
85
86 Total Reward Received: 1
87 Taking a look at model parameters to see if
    weights are changing
88 Parameter containing:
89 tensor([[ 0.0361],
90          [-0.0237],
91          [-0.0058]], dtype=torch.float64,
92 Parameter containing:
93 tensor([0.1189, 0.1003, 0.1022], dtype=torch.
94 Starting Training
95 layer weights: Parameter containing:
96 tensor([[-0.8989],
97          [ 0.6057],
98          [-0.5884]], dtype=torch.float64,
99 layer weights: Parameter containing:
100 tensor([[-0.5352],
101          [ 0.2788],
102          [ 0.0697]], dtype=torch.float64,
103 Summary AFTER training
104
105 =====
106 Summary:
107 State  Action  Next State  Reward
108 0      0      Q(s,a)= tensor(0.1288, dtype=
109 0      1      Q(s,a)= tensor(0.1023, dtype=
           torch.float64, grad_fn=<SelectBackward0>)
           torch.float64, grad_fn=<SelectBackward0>)
```

```

110 0      2      Q(s,a)= tensor(0.1052, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
111 1      0      Q(s,a)= tensor(0.1752, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
112 1      1      Q(s,a)= tensor(0.0829, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
113 1      2      Q(s,a)= tensor(0.1001, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
114 2      0      Q(s,a)= tensor(0.2216, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
115 2      1      Q(s,a)= tensor(0.0635, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
116 2      2      Q(s,a)= tensor(0.0951, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
117 3      0      Q(s,a)= tensor(0.2681, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
118 3      1      Q(s,a)= tensor(0.0441, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
119 3      2      Q(s,a)= tensor(0.0900, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
120 4      0      Q(s,a)= tensor(0.3145, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
121 4      1      Q(s,a)= tensor(0.0247, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
122 4      2      Q(s,a)= tensor(0.0849, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
123 =====
124
125 Best trajectory:
126 Best trajectory from Test Environment
127 State: tensor([0.], dtype=torch.float64) Action
    : 0 Reward Received: 0
128 State: tensor([1.], dtype=torch.float64) Action
    : 0 Reward Received: 0
129 State: tensor([2.], dtype=torch.float64) Action
    : 0 Reward Received: 0
130 State: tensor([3.], dtype=torch.float64) Action
    : 0 Reward Received: 1
131
132 Total Reward Received: 1
133 Taking a look at model parameters to see if

```

```
133 weights are changing
134 Parameter containing:
135 tensor([[ 0.0464],
136             [-0.0194],
137             [-0.0051]], dtype=torch.float64,
138             requires_grad=True)
138 Parameter containing:
139 tensor([0.1288, 0.1023, 0.1052], dtype=torch.
140             float64, requires_grad=True)
140 Starting Training
141 layer weights: Parameter containing:
142 tensor([[ 0.9269],
143             [-0.1591],
144             [ 0.7415]], dtype=torch.float64,
145             requires_grad=True)
145 layer weights: Parameter containing:
146 tensor([[-0.8975],
147             [ 0.5026],
148             [ 0.7971]], dtype=torch.float64,
149             requires_grad=True)
149 Summary AFTER training
150
151 =====
152 Summary:
153 State    Action    Next State    Reward
154 0        0        Q(s,a)= tensor(0.1077, dtype=
155             torch.float64, grad_fn=<SelectBackward0>)
155 0        1        Q(s,a)= tensor(0.0882, dtype=
156             torch.float64, grad_fn=<SelectBackward0>)
156 0        2        Q(s,a)= tensor(0.0905, dtype=
157             torch.float64, grad_fn=<SelectBackward0>)
157 1        0        Q(s,a)= tensor(0.1454, dtype=
158             torch.float64, grad_fn=<SelectBackward0>)
158 1        1        Q(s,a)= tensor(0.0728, dtype=
159             torch.float64, grad_fn=<SelectBackward0>)
159 1        2        Q(s,a)= tensor(0.0866, dtype=
160             torch.float64, grad_fn=<SelectBackward0>)
160 2        0        Q(s,a)= tensor(0.1832, dtype=
161             torch.float64, grad_fn=<SelectBackward0>)
161 2        1        Q(s,a)= tensor(0.0575, dtype=
```

```
162 2          2      Q(s,a)= tensor(0.0826, dtype=
163   torch.float64, grad_fn=<SelectBackward0>)
164 3          0      Q(s,a)= tensor(0.2209, dtype=
165   torch.float64, grad_fn=<SelectBackward0>)
166 3          1      Q(s,a)= tensor(0.0421, dtype=
167   torch.float64, grad_fn=<SelectBackward0>)
168 3          2      Q(s,a)= tensor(0.0787, dtype=
169   torch.float64, grad_fn=<SelectBackward0>)
170 4          0      Q(s,a)= tensor(0.2587, dtype=
171   torch.float64, grad_fn=<SelectBackward0>)
172 4          1      Q(s,a)= tensor(0.0267, dtype=
173   torch.float64, grad_fn=<SelectBackward0>)
174 4          2      Q(s,a)= tensor(0.0748, dtype=
175   torch.float64, grad_fn=<SelectBackward0>)
176 =====
177
178 Best trajectory:
179 Best trajectory from Test Environment
180 State: tensor([0.], dtype=torch.float64) Action
181 : 0 Reward Received: 0
182 State: tensor([1.], dtype=torch.float64) Action
183 : 0 Reward Received: 0
184 State: tensor([2.], dtype=torch.float64) Action
185 : 0 Reward Received: 0
186 State: tensor([3.], dtype=torch.float64) Action
187 : 0 Reward Received: 1
188
189 Total Reward Received: 1
190 Taking a look at model parameters to see if
191 weights are changing
192 Parameter containing:
193 tensor([[ 0.0377],
194         [-0.0154],
195         [-0.0039]], dtype=torch.float64,
196         requires_grad=True)
197 Parameter containing:
198 tensor([0.1077, 0.0882, 0.0905], dtype=torch.
199         float64, requires_grad=True)
200 Starting Training
201 layer weights: Parameter containing:
202 tensor([[0.2033],
```

```
189      [0.0861],  
190      [0.6052]], dtype=torch.float64,  
    requires_grad=True)  
191 layer weights: Parameter containing:  
192 tensor([[ 0.1543],  
193           [ 0.8613],  
194           [-0.7033]], dtype=torch.float64,  
    requires_grad=True)  
195 Summary AFTER training  
196  
197 =====  
198 Summary:  
199 State  Action  Next State  Reward  
200 0      0      Q(s,a)= tensor(0.1549, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
201 0      1      Q(s,a)= tensor(0.1230, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
202 0      2      Q(s,a)= tensor(0.1274, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
203 1      0      Q(s,a)= tensor(0.2062, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
204 1      1      Q(s,a)= tensor(0.1003, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
205 1      2      Q(s,a)= tensor(0.1152, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
206 2      0      Q(s,a)= tensor(0.2576, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
207 2      1      Q(s,a)= tensor(0.0776, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
208 2      2      Q(s,a)= tensor(0.1031, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
209 3      0      Q(s,a)= tensor(0.3090, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
210 3      1      Q(s,a)= tensor(0.0549, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
211 3      2      Q(s,a)= tensor(0.0909, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
212 4      0      Q(s,a)= tensor(0.3603, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)  
213 4      1      Q(s,a)= tensor(0.0322, dtype=  
    torch.float64, grad_fn=<SelectBackward0>)
```

```
214 4          2          Q(s,a)= tensor(0.0787, dtype=
215                               torch.float64, grad_fn=<SelectBackward0>)
216 =====
217 Best trajectory:
218 Best trajectory from Test Environment
219 State: tensor([0.], dtype=torch.float64) Action
220 : 0 Reward Received: 0
221 State: tensor([1.], dtype=torch.float64) Action
222 : 0 Reward Received: 0
223 State: tensor([2.], dtype=torch.float64) Action
224 : 0 Reward Received: 0
225 State: tensor([3.], dtype=torch.float64) Action
226 : 0 Reward Received: 1
227
228 Total Reward Received: 1
229 Taking a look at model parameters to see if
230 weights are changing
231 Parameter containing:
232 tensor([[ 0.0514],
233           [-0.0227],
234           [-0.0122]], dtype=torch.float64,
235           requires_grad=True)
236 Parameter containing:
237 tensor([0.1549, 0.1230, 0.1274], dtype=torch.
238 float64, requires_grad=True)
239 Starting Training
240 layer weights: Parameter containing:
241 tensor([[-0.8031],
242           [-0.6806],
243           [ 0.4432]], dtype=torch.float64,
244           requires_grad=True)
245 layer weights: Parameter containing:
246 tensor([[ 0.7594],
247           [ 0.0182],
248           [-0.9328]], dtype=torch.float64,
249           requires_grad=True)
250 Summary AFTER training
251
252 =====
253 Summary:
```

```

245 State    Action   Next State   Reward
246 0          0        Q(s,a)= tensor(0.1310, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
247 0          1        Q(s,a)= tensor(0.1118, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
248 0          2        Q(s,a)= tensor(0.1108, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
249 1          0        Q(s,a)= tensor(0.1644, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
250 1          1        Q(s,a)= tensor(0.0867, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
251 1          2        Q(s,a)= tensor(0.0977, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
252 2          0        Q(s,a)= tensor(0.1977, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
253 2          1        Q(s,a)= tensor(0.0616, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
254 2          2        Q(s,a)= tensor(0.0845, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
255 3          0        Q(s,a)= tensor(0.2311, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
256 3          1        Q(s,a)= tensor(0.0366, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
257 3          2        Q(s,a)= tensor(0.0714, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
258 4          0        Q(s,a)= tensor(0.2645, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
259 4          1        Q(s,a)= tensor(0.0115, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
260 4          2        Q(s,a)= tensor(0.0583, dtype=
      torch.float64, grad_fn=<SelectBackward0>)
261 =====
262
263 Best trajectory:
264 Best trajectory from Test Environment
265 State: tensor([0.], dtype=torch.float64) Action
      : 0 Reward Received: 0
266 State: tensor([1.], dtype=torch.float64) Action
      : 0 Reward Received: 0
267 State: tensor([2.], dtype=torch.float64) Action
      : 0 Reward Received: 0

```

```
268 State: tensor([3.], dtype=torch.float64) Action
    : 0 Reward Received: 1
269
270 Total Reward Received: 1
271 Taking a look at model parameters to see if
    weights are changing
272 Parameter containing:
273 tensor([[ 0.0334],
274           [-0.0251],
275           [-0.0131]], dtype=torch.float64,
    requires_grad=True)
276 Parameter containing:
277 tensor([0.1310, 0.1118, 0.1108], dtype=torch.
    float64, requires_grad=True)
278 Starting Training
279 layer weights: Parameter containing:
280 tensor([[ 0.1839],
281           [ 0.1137],
282           [-0.2649]], dtype=torch.float64,
    requires_grad=True)
283 layer weights: Parameter containing:
284 tensor([[-0.4017],
285           [ 0.7819],
286           [ 0.7345]], dtype=torch.float64,
    requires_grad=True)
287 Summary AFTER training
288
289 =====
290 Summary:
291 State  Action  Next State  Reward
292 0      0      Q(s,a)= tensor(0.1459, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
293 0      1      Q(s,a)= tensor(0.1178, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
294 0      2      Q(s,a)= tensor(0.1205, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
295 1      0      Q(s,a)= tensor(0.1927, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
296 1      1      Q(s,a)= tensor(0.0958, dtype=
    torch.float64, grad_fn=<SelectBackward0>)
297 1      2      Q(s,a)= tensor(0.1160, dtype=
```

```
297 torch.float64, grad_fn=<SelectBackward0>)
298 2          0      Q(s,a)= tensor(0.2394, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
299 2          1      Q(s,a)= tensor(0.0737, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
300 2          2      Q(s,a)= tensor(0.1115, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
301 3          0      Q(s,a)= tensor(0.2862, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
302 3          1      Q(s,a)= tensor(0.0517, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
303 3          2      Q(s,a)= tensor(0.1070, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
304 4          0      Q(s,a)= tensor(0.3329, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
305 4          1      Q(s,a)= tensor(0.0296, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
306 4          2      Q(s,a)= tensor(0.1025, dtype=
   torch.float64, grad_fn=<SelectBackward0>)
307 =====
308
309 Best trajectory:
310 Best trajectory from Test Environment
311 State: tensor([0.], dtype=torch.float64) Action
   : 0 Reward Received: 0
312 State: tensor([1.], dtype=torch.float64) Action
   : 0 Reward Received: 0
313 State: tensor([2.], dtype=torch.float64) Action
   : 0 Reward Received: 0
314 State: tensor([3.], dtype=torch.float64) Action
   : 0 Reward Received: 1
315
316 Total Reward Received: 1
317 Taking a look at model parameters to see if
   weights are changing
318 Parameter containing:
319 tensor([[ 0.0467],
320           [-0.0220],
321           [-0.0045]], dtype=torch.float64,
   requires_grad=True)
322 Parameter containing:
```

```
323 tensor([0.1459, 0.1178, 0.1205], dtype=torch.  
          float64, requires_grad=True)  
324 Starting Training  
325 layer weights: Parameter containing:  
326 tensor([[ 0.0635],  
327             [ 0.0842],  
328             [-0.2756]], dtype=torch.float64,  
            requires_grad=True)  
329 layer weights: Parameter containing:  
330 tensor([[0.3867],  
331             [0.5861],  
332             [0.7759]], dtype=torch.float64,  
            requires_grad=True)  
333 Summary AFTER training  
334  
335 =====  
336 Summary:  
337 State    Action   Next State   Reward  
338 0         0        Q(s,a)= tensor(0.1306, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
339 0         1        Q(s,a)= tensor(0.1096, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
340 0         2        Q(s,a)= tensor(0.1104, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
341 1         0        Q(s,a)= tensor(0.1798, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
342 1         1        Q(s,a)= tensor(0.0911, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
343 1         2        Q(s,a)= tensor(0.1053, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
344 2         0        Q(s,a)= tensor(0.2289, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
345 2         1        Q(s,a)= tensor(0.0725, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
346 2         2        Q(s,a)= tensor(0.1003, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
347 3         0        Q(s,a)= tensor(0.2781, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
348 3         1        Q(s,a)= tensor(0.0539, dtype=  
           torch.float64, grad_fn=<SelectBackward0>)  
349 3         2        Q(s,a)= tensor(0.0952, dtype=
```

```
349 torch.float64, grad_fn=<SelectBackward0>)
350 4          0          Q(s,a)= tensor(0.3273, dtype=
351               torch.float64, grad_fn=<SelectBackward0>)
352 4          1          Q(s,a)= tensor(0.0353, dtype=
353               torch.float64, grad_fn=<SelectBackward0>)
354 4          2          Q(s,a)= tensor(0.0902, dtype=
355               torch.float64, grad_fn=<SelectBackward0>)
356 =====
357
358 Best trajectory:
359 Best trajectory from Test Environment
360 State: tensor([0.], dtype=torch.float64) Action
361 : 0 Reward Received: 0
362 State: tensor([1.], dtype=torch.float64) Action
363 : 0 Reward Received: 0
364 State: tensor([2.], dtype=torch.float64) Action
365 : 0 Reward Received: 0
366 State: tensor([3.], dtype=torch.float64) Action
367 : 0 Reward Received: 1
368
369 Total Reward Received: 1
370 Taking a look at model parameters to see if
371 weights are changing
372 Parameter containing:
373 tensor([[ 0.0492],
374         [-0.0186],
375         [-0.0050]], dtype=torch.float64,
376         requires_grad=True)
377 Parameter containing:
378 tensor([0.1306, 0.1096, 0.1104], dtype=torch.
379 float64, requires_grad=True)
380 Starting Training
381 layer weights: Parameter containing:
382 tensor([[ 0.2973],
383         [ 0.5562],
384         [-0.6050]], dtype=torch.float64,
385         requires_grad=True)
386 layer weights: Parameter containing:
387 tensor([[-0.0488],
388         [-0.5204],
389         [-0.0330]], dtype=torch.float64,
```

```
378 requires_grad=True)
379 Summary AFTER training
380
381 =====
382 Summary:
383 State    Action   Next State   Reward
384 0          0        Q(s,a)= tensor(0.1317, dtype=
torch.float64, grad_fn=<SelectBackward0>)
385 0          1        Q(s,a)= tensor(0.1079, dtype=
torch.float64, grad_fn=<SelectBackward0>)
386 0          2        Q(s,a)= tensor(0.1088, dtype=
torch.float64, grad_fn=<SelectBackward0>)
387 1          0        Q(s,a)= tensor(0.1741, dtype=
torch.float64, grad_fn=<SelectBackward0>)
388 1          1        Q(s,a)= tensor(0.0862, dtype=
torch.float64, grad_fn=<SelectBackward0>)
389 1          2        Q(s,a)= tensor(0.1042, dtype=
torch.float64, grad_fn=<SelectBackward0>)
390 2          0        Q(s,a)= tensor(0.2165, dtype=
torch.float64, grad_fn=<SelectBackward0>)
391 2          1        Q(s,a)= tensor(0.0644, dtype=
torch.float64, grad_fn=<SelectBackward0>)
392 2          2        Q(s,a)= tensor(0.0996, dtype=
torch.float64, grad_fn=<SelectBackward0>)
393 3          0        Q(s,a)= tensor(0.2589, dtype=
torch.float64, grad_fn=<SelectBackward0>)
394 3          1        Q(s,a)= tensor(0.0426, dtype=
torch.float64, grad_fn=<SelectBackward0>)
395 3          2        Q(s,a)= tensor(0.0950, dtype=
torch.float64, grad_fn=<SelectBackward0>)
396 4          0        Q(s,a)= tensor(0.3013, dtype=
torch.float64, grad_fn=<SelectBackward0>)
397 4          1        Q(s,a)= tensor(0.0209, dtype=
torch.float64, grad_fn=<SelectBackward0>)
398 4          2        Q(s,a)= tensor(0.0904, dtype=
torch.float64, grad_fn=<SelectBackward0>)
399 =====
400
401 Best trajectory:
402 Best trajectory from Test Environment
403 State: tensor([0.], dtype=torch.float64)  Action
```

```
403 : 0 Reward Received: 0
404 State: tensor([1.], dtype=torch.float64) Action
        : 0 Reward Received: 0
405 State: tensor([2.], dtype=torch.float64) Action
        : 0 Reward Received: 0
406 State: tensor([3.], dtype=torch.float64) Action
        : 0 Reward Received: 1
407
408 Total Reward Received: 1
409 Taking a look at model parameters to see if
    weights are changing
410 Parameter containing:
411 tensor([[ 0.0424],
412           [-0.0218],
413           [-0.0046]], dtype=torch.float64,
414         requires_grad=True)
414 Parameter containing:
415 tensor([0.1317, 0.1079, 0.1088], dtype=torch.
416         float64, requires_grad=True)
416 Starting Training
417 layer weights: Parameter containing:
418 tensor([[-0.2418],
419           [-0.2701],
420           [ 0.3177]], dtype=torch.float64,
421         requires_grad=True)
421 layer weights: Parameter containing:
422 tensor([[-0.7138],
423           [-0.0133],
424           [-0.2739]], dtype=torch.float64,
425         requires_grad=True)
425 Summary AFTER training
426
427 =====
428 Summary:
429 State  Action  Next State  Reward
430 0      0      Q(s,a)= tensor(0.1413, dtype=
431         torch.float64, grad_fn=<SelectBackward0>)
431 0      1      Q(s,a)= tensor(0.1149, dtype=
432         torch.float64, grad_fn=<SelectBackward0>)
432 0      2      Q(s,a)= tensor(0.1166, dtype=
433         torch.float64, grad_fn=<SelectBackward0>)
```

```

433 1          0      Q(s,a)= tensor(0.1904, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
434 1          1      Q(s,a)= tensor(0.0940, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
435 1          2      Q(s,a)= tensor(0.1050, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
436 2          0      Q(s,a)= tensor(0.2395, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
437 2          1      Q(s,a)= tensor(0.0731, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
438 2          2      Q(s,a)= tensor(0.0935, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
439 3          0      Q(s,a)= tensor(0.2886, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
440 3          1      Q(s,a)= tensor(0.0523, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
441 3          2      Q(s,a)= tensor(0.0820, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
442 4          0      Q(s,a)= tensor(0.3377, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
443 4          1      Q(s,a)= tensor(0.0314, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
444 4          2      Q(s,a)= tensor(0.0705, dtype=
        torch.float64, grad_fn=<SelectBackward0>)
445 =====
446
447 Best trajectory:
448 Best trajectory from Test Environment
449 State: tensor([0.], dtype=torch.float64) Action
        : 0 Reward Received: 0
450 State: tensor([1.], dtype=torch.float64) Action
        : 0 Reward Received: 0
451 State: tensor([2.], dtype=torch.float64) Action
        : 0 Reward Received: 0
452 State: tensor([3.], dtype=torch.float64) Action
        : 0 Reward Received: 1
453
454 Total Reward Received: 1
455 Taking a look at model parameters to see if
        weights are changing
456 Parameter containing:

```

```
457 tensor([[ 0.0491],  
458             [-0.0209],  
459             [-0.0115]], dtype=torch.float64,  
460             requires_grad=True)  
460 Parameter containing:  
461 tensor([0.1413, 0.1149, 0.1166], dtype=torch.  
462             float64, requires_grad=True)  
462  
463 Process finished with exit code 0  
464
```