


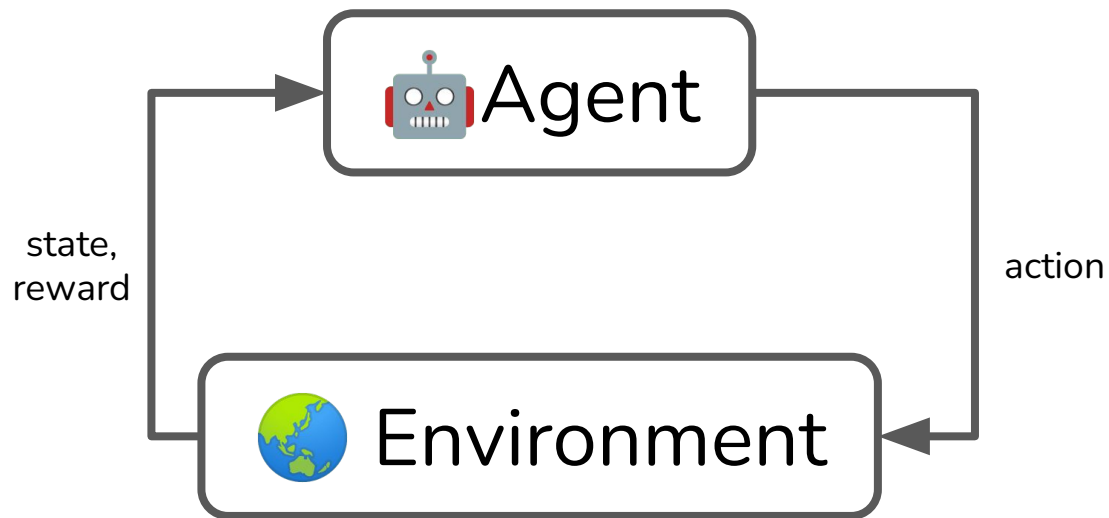
Learning Reinforcement Learning with OpenAI Gym

Ali Akbar Septiandri
@aliakbars







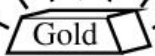










Did you get into IT
because you wanted
to be a game dev? 

Anyone familiar with this image?

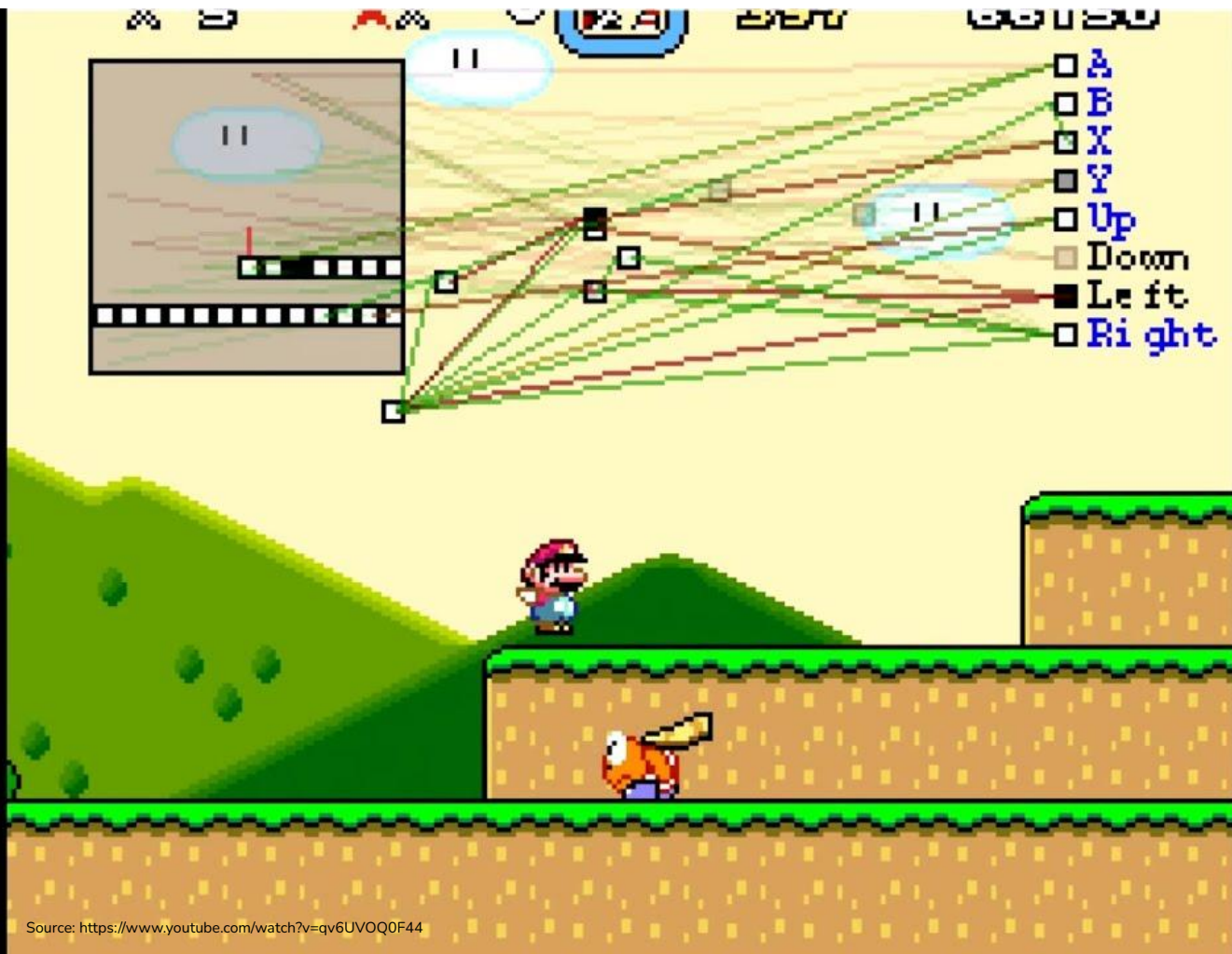


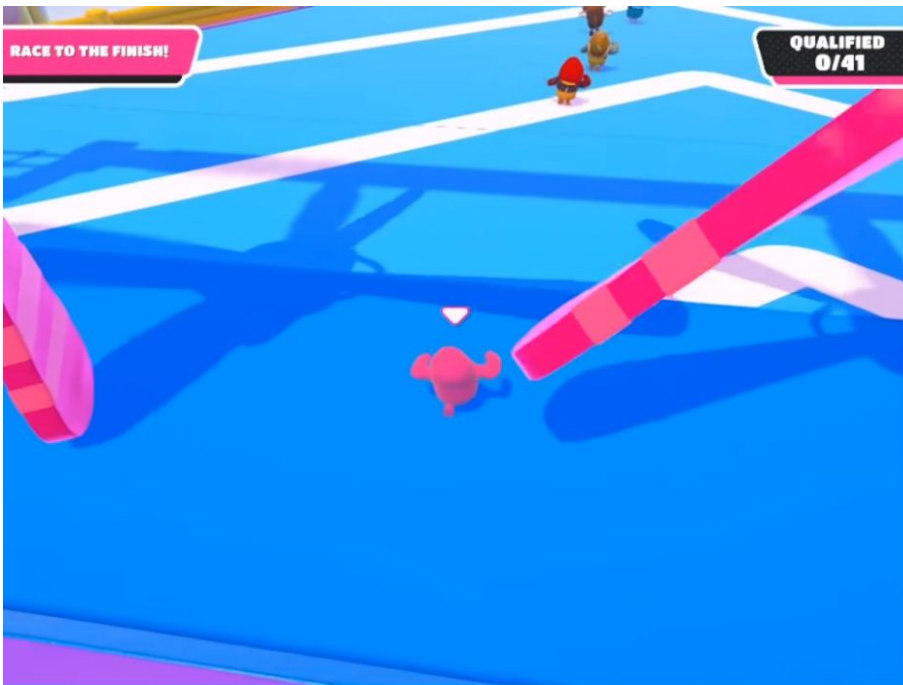
What about this one?

 Stench		 Breeze	 PIT
	 Breeze  Stench  Gold	 PIT	 Breeze
 Stench		 Breeze	
 START	 Breeze	 PIT	 Breeze

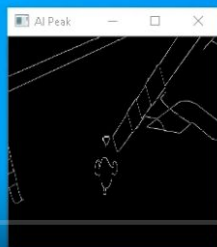
$$V_{\pi}^{(t)}(s) \leftarrow \sum_{s'} T(s, \pi(s), s') [R(s, \pi(s), s') + \gamma V_{\pi}^{(t-1)}(s')]$$

What about this? 





JUMP!



1:53 / 12:12

Source: https://www.youtube.com/watch?v=GS_0ZKzrvk0

```
File Edit Selection View Go Run Terminal Help CreateData.py - FallGuys - Visual Studio Code
TrainedAgent.py CreateData.py x Settings
CreateData.py > ...
1 import numpy as np
2 import cv2
3 import time
4 import os
5
6 from utils.grabscreen import grab_screen
7 from utils.getkeys import key_check
8
9
10 file_name = "C:/Users/programmer/Desktop/FallGuys/data/training"
11 file_name2 = "C:/Users/programmer/Desktop/FallGuys/data/target"
12
13
14 def auto_canny(image, sigma=0.33):
15     # compute the median of the single channel pixel intensities
16     v = np.median(image)
17     # apply automatic Canny edge detection using the computed median
18     lower = int(max(0, (1.0 - sigma) * v))
19     print(lower)
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

OUTPUT TERMINAL DEBUG CONSOLE PROBLEMS 0 2: Python

```
loop took 0.031914718998535156 seconds
loop took 0.032911862248680596 seconds
loop took 0.03391146659851974 seconds
loop took 0.03391083688703613 seconds
loop took 0.03251298221358099 seconds
loop took 0.03391122817993164 seconds
loop took 0.03291177749633789 seconds
loop took 0.032910823822821484 seconds
loop took 0.03091573715289961 seconds
loop took 0.03391075134277344 seconds
loop took 0.03191389574279785 seconds
loop took 0.0219418882532959 seconds
loop took 0.03091716766357422 seconds
loop took 0.03598758694274982 seconds
loop took 0.029922862188728783 seconds
loop took 0.0358630783357129 seconds
loop took 0.030918121337898625 seconds
loop took 0.03490614891852246 seconds
```

Python 3.8.3 64-bit

ALPHAGO
00:05:30



 Google DeepMind
Challenge Match



LEE SEDOL
00:28:28



**LULUS S1
MAU JADI
SOFTWARE ENGINEER**

**LULUS S1
MAU JADI
DATA SCIENTIST**

**LULUS S1 MAU
JADI RESEARCH
SCIENTIST DI GOOGLE**

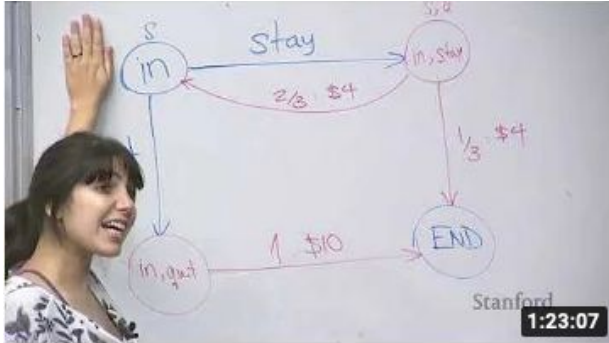
**LULUS S1 MAU
BIKIN ARTIFICIAL
GENERAL INTELLIGENCE**



A Whirlwind Tour of Reinforcement Learning



Stanford CS221: Artificial Intelligence



Lecture 7: Markov Decision Processes - Value Iteration | Stanford CS221: AI (Autumn 2019)

91K views · 1 year ago



For more information about Stanford's Artificial Intelligence professional and graduate programs, visit: <https://stanford.io/3pUNqG7> ...

8:08 Um, and in the middle, I'm going to talk about policy evaluation, which is not an inference algorithm but it's kind of a step toward...

Subtitles



Lecture 8: Markov Decision Processes - Reinforcement Learning | Stanford CS221: AI (Autumn 2019)

26K views · 1 year ago



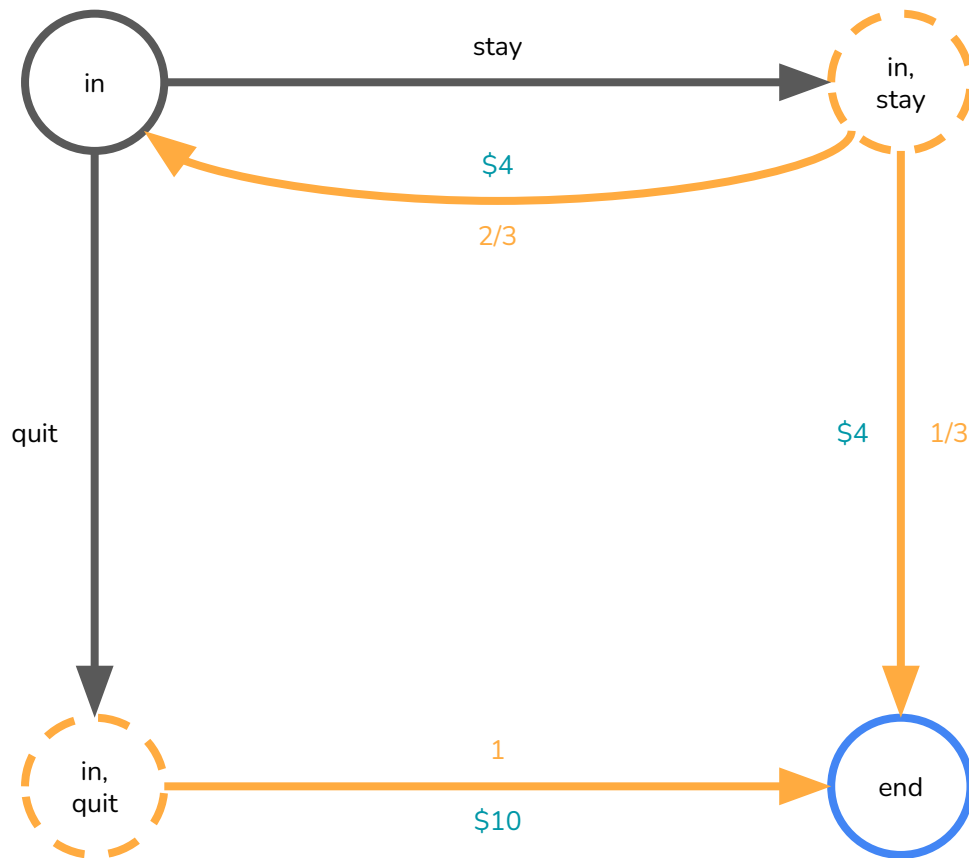
For more information about Stanford's Artificial Intelligence professional and graduate programs, visit: <https://stanford.io/2Zv1JpK> ...

49:00 So in MDPs, we saw that policy evaluation allows you to get Q Pi; value iteration get- allows you to get Q opt. And now, we're ...

Subtitles

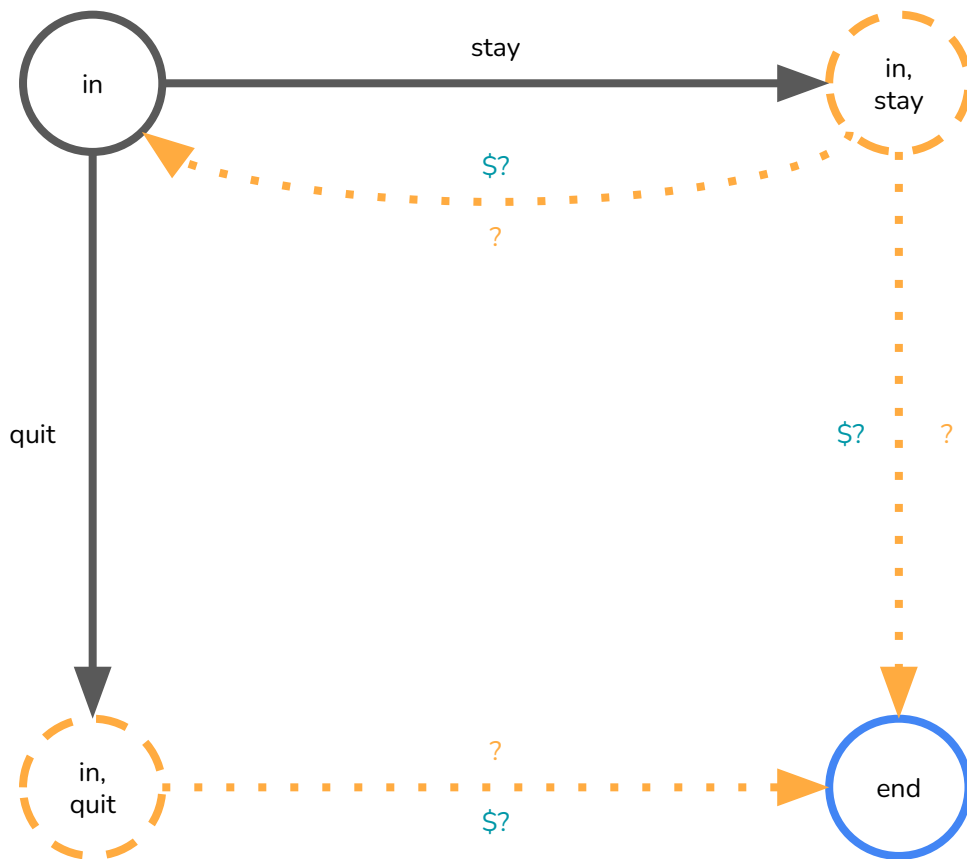
Markov Decision Process

- States
- $s_{\text{start}} \in \text{States}$
- Actions(s)
- $T(s, a, s')$
- $\text{Reward}(s, a, s')$
- $\text{IsEnd}(s)$
- $0 \leq \gamma \leq 1$



Reinforcement Learning

- States
- $s_{\text{start}} \in \text{States}$
- $\text{Actions}(s)$
-
-
- $\text{IsEnd}(s)$
- $0 \leq \gamma \leq 1$



$$V_{\pi}^{(t)}(s) \leftarrow \sum_{s'} T(s, \pi(s), s') [R(s, \pi(s), s') + \gamma V_{\pi}^{(t-1)}(s')]$$

Model-based

Monte Carlo



Do we have to explore
all states?

SARSA &
Q-Learning



Evaluation vs Iteration

SARSA

On each (s, a, r, s', a') :

$$\hat{Q}_\pi(s, a) \leftarrow (1 - \eta)\hat{Q}_\pi(s, a) + \eta(r + \gamma\hat{Q}_\pi(s', a'))$$

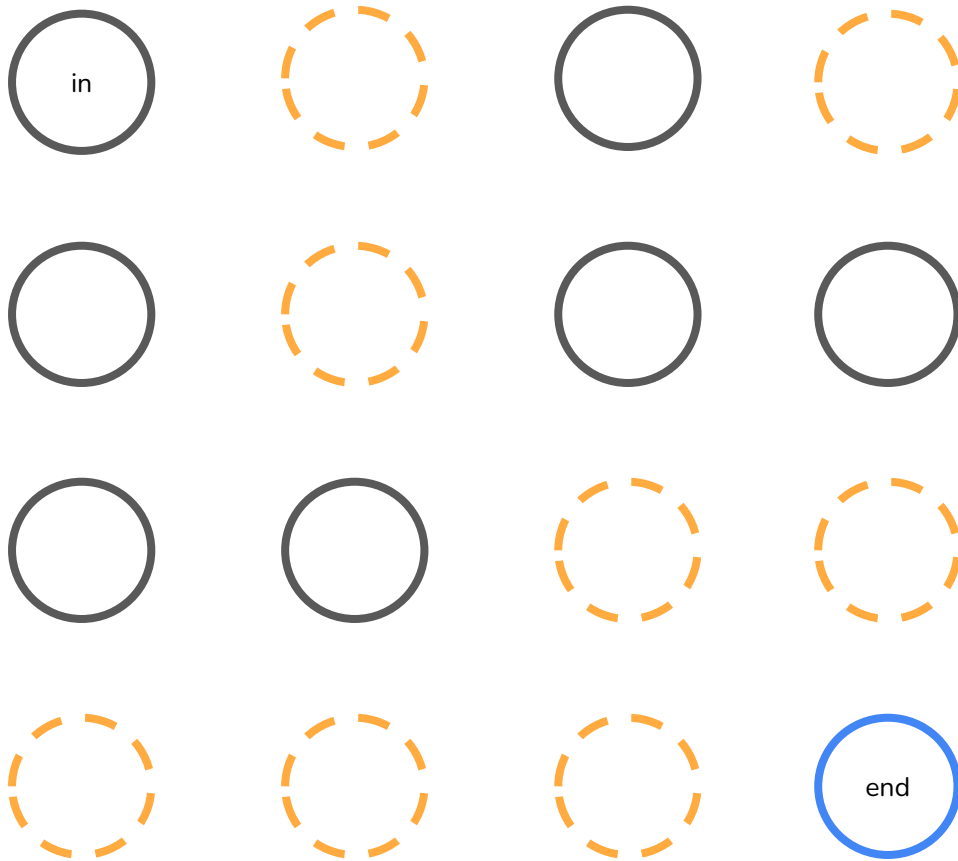
Q-learning

On each (s, a, r, s') :

$$\hat{Q}_{\text{opt}}(s, a) \leftarrow (1 - \eta)\hat{Q}_{\text{opt}}(s, a) + \eta(r + \gamma \max_{a' \in \text{Actions}(s')} \hat{Q}_{\text{opt}}(s', a'))]$$

Reinforcement Learning

- States
- $s_{\text{start}} \in \text{States}$
- Actions(s)
-
-
- IsEnd(s)
- $0 \leq \gamma \leq 1$



Unlike classical ML,
you don't really have
scikit-learn and
ImageNet / Kaggle

OpenAI Gym



Gym

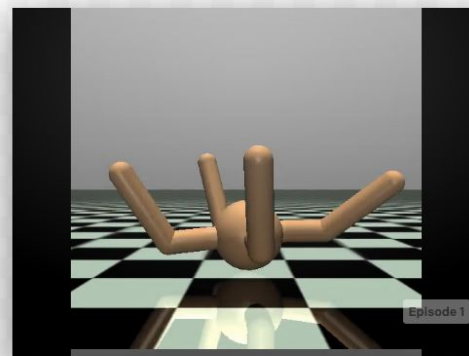
Gym is a toolkit for developing and comparing reinforcement learning algorithms. It supports teaching agents everything from walking to playing games like Pong or Pinball.

[View documentation >](#)

[View on GitHub >](#)



RandomAgent on LunarLander-v2



RandomAgent on Ant-v2



Algorithms

Atari

Box2D

Classic control

MuJoCo

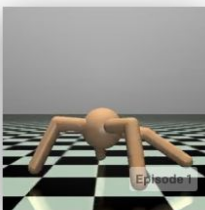
Robotics

Toy text EASY

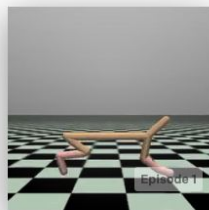
Third party environments [↗](#)

MuJoCo

Continuous control tasks, running in a fast physics simulator.



Ant-v2
Make a 3D four-legged robot walk.



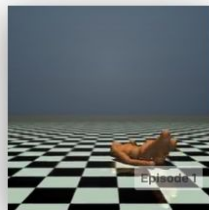
HalfCheetah-v2
Make a 2D cheetah robot run.



Hopper-v2
Make a 2D robot hop.



Humanoid-v2
Make a 3D two-legged robot walk.

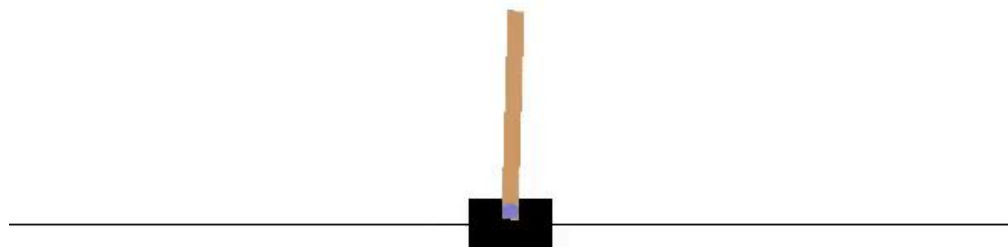


HumanoidStandup-v2
Make a 3D two-legged robot stand up.



InvertedDoublePendulum-v2
Balance a pole on a pole on

Examples





```
import gym
env = gym.make('CartPole-v0')
for i_episode in range(20):
    observation = env.reset()
    for t in range(100):
        env.render()
        print(observation)
        action = env.action_space.sample()
        observation, reward, done, info = env.step(action)
        if done:
            print("Episode finished after {} timesteps".format(t+1))
            break
env.close()
```

Demo

Evaluation vs Iteration

SARSA

On each (s, a, r, s', a') :

$$\hat{Q}_\pi(s, a) \leftarrow (1 - \eta)\hat{Q}_\pi(s, a) + \eta(r + \gamma\hat{Q}_\pi(s', a'))$$

Q-learning

On each (s, a, r, s') :

$$\hat{Q}_{\text{opt}}(s, a) \leftarrow (1 - \eta)\hat{Q}_{\text{opt}}(s, a) + \eta(r + \gamma \max_{a' \in \text{Actions}(s')} \hat{Q}_{\text{opt}}(s', a'))]$$

Function
approximation



$$\hat{Q}_{\text{opt}}(s, a; \mathbf{w}) = \mathbf{w} \cdot \phi(s, a)$$




Q can be approximated as weights x features

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \left[\underbrace{\hat{Q}_{\text{opt}}(s, a; \mathbf{w})}_{\text{prediction}} - \underbrace{(r + \gamma \hat{V}_{\text{opt}}(s'))}_{\text{target}} \right] \phi(s, a)$$

On each (s, a, r, s') , fix the weights

Combining with Deep
Learning 

RL Applications

- Game playing 
- Multi-armed bandits 
- Dialogue systems 
- ...

Learning Resources

1. [Stanford CS221: Artificial Intelligence](#)
2. [DeepMind/UCL: Reinforcement Learning](#) by David Silver
3. [UC Berkeley CS285: Deep Reinforcement Learning](#) by Sergey Levine
4. [Reinforcement Learning: An Introduction](#) by Sutton & Barto

Conclusions

1. RL is often overlooked in ML syllabi because of the hassles 🙄
2. OpenAI Gym can help you teach/learn RL – from Q-learning to DQN! 🏋️💪
3. Learning by doing is the best 🧑💻🐱💻🧑💻

Thank you

@aliakbars