NOTE BOOK

Rokshana Ahmed

DEEP LEARNING

Date: 30 / 11 / 2023

Previously on Deep Learning N

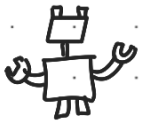
"Previously, on Breaking Bad..."
cit. Tony

NOTE: 1st time, we can see a wild Tony with no beanie

Unsupervised Learning (Unlabelled data)

- Generative Adversal Networks (GANs)
- Restricted Boltzman Machines (RBMs)
- Transformers (later)
  - ChatGPT
  - Dall E

Today: DEEP REINFORCEMENT LEARNING (no data)

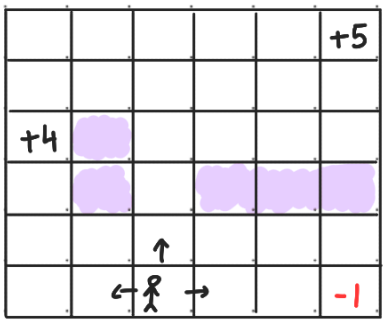
Agent =  = SUPER MARIO

Environment = VIDEOGAME

Agent learns by interacting with the environment

Agent = investor
Environment = Trading platform

GRIDWORLD



RULES

- moves around the grid
- Game is over when he lands on one of those 3 squares.
- He collects the points on that square.

State = □ (called s)

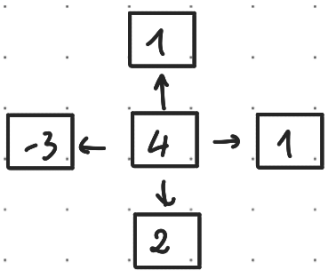
Actions = ↑, ↓, ←, → (called a)

DEEP LEARNING

Date: 30/ 11 /2023

VALUE : $\boxed{V(s) = \max_a \left(V(s') \right)}$

where s' is the state reachable from s by taking action a



5	5	5	5	5	+5
5	5	5	5	5	5
+4		5	5	5	5
5		5			
5	5	5	5	5	5
5	5	5	5	5	-1

POLICY = a set of actions for each state

$\boxed{\text{always move to the square with the highest value}}$

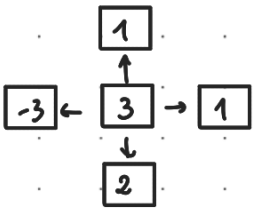
5	5	5	5	5	+5
5	5	5	5	5	5
+4		5	5	5	5
5		5			
5	5	5	5	5	5
5	5	5	5	5	-1

Reward $R(s,a)$ — s: state, a: action

can be positive or negative

For us, let's say $R(s,a)$ is always -1

Value (updated) $\boxed{V(s) = \max_a \left(R(s,a) + V(s') \right)}$



2	1	2	3	4	+5
3	2	1	2	3	4
+4		0	1	2	3
3		-1			
2	1	0	-1	-2	-2
1	0	-1	-2	-2	-1

↓	←↓→	→	→	→	+5
↓	←	←↑→	↑→	↑→	↑
+4		↑→	↑→	↑→	↑
↑		↑↓			
↑	←	←	←	←	↓
↑	↑←	↑←	↑←	→	-1

policy

DEEP LEARNING

Date: 30 / 11 / 2023

DISCOUNT FACTOR γ

TODAY

$\parallel$
$1

TOMORROW

$\parallel$
$ 0.90

TWO DAYS FROM NOW

$\parallel$
$ 0.81

3.24	2.95	3.28	4.05	4.5	+5
3.6	3.24	2.95	3.28	4.05	4.5
+4		2.65	2.95	3.28	4.05
3.6		2.39			
3.24	2.92	2.62	2.36	2.13	1.91
2.92	2.62	2.36	2.13	1.91	-1

↓	→	→	→	→	+5
↓	←	↑→	↑→	↑→	↑
+4		↑→	↑→	↑→	↑
↑		↑			
↑→	←	←	←	←	←
↑	↑←	↑←	↑←	↑	-1

VALUE
(updated.... again)

$$\boxed{V(s) = \max_a \left(R(s,a) + \gamma V(s')\right)}$$

s' = state obtained from s by performing a

$R(s,a)$ = reward for performing "a" at state s

γ = discount factor

Sometimes there is a probability term
(today + sometime)

HOW DO WE COMPUTE THE VALUES?
(Solving the Bellman 🔔 equation)

Can't really feel my hand anymore ...
RIP me T.T

1. Choose a random state S
   Fill in a value for s satisfying the Bellman equation
   (If no state adjacent to s has a value, do nothing)

2. Repeat many times.

"Wonder how much long should i continue.... Mmh i can stop."
- cit Tony at 15:40
(class ends at 16)

⇒ PROCEEDS TO CONTINUE
(it's ok, he cool)

THERE IS A PROBLEM!

Typically there are too many states and actions to compute the values exactly.

Instead we will use a NN to find an approximate solution to the Bellman Equation