Davide Basso - *UniTS & Infineon Technologies PhD Student*

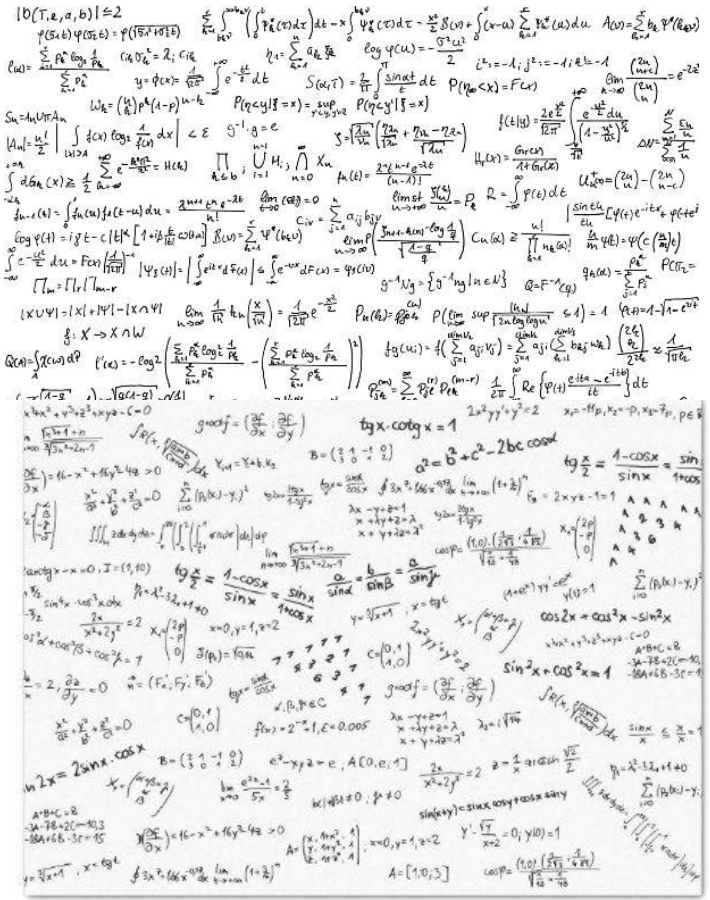
REINFORCEMENT LEARNING

Imitando l'apprendimento umano

UNIVERSITÀ DEGLI STUDI DI TRIESTE

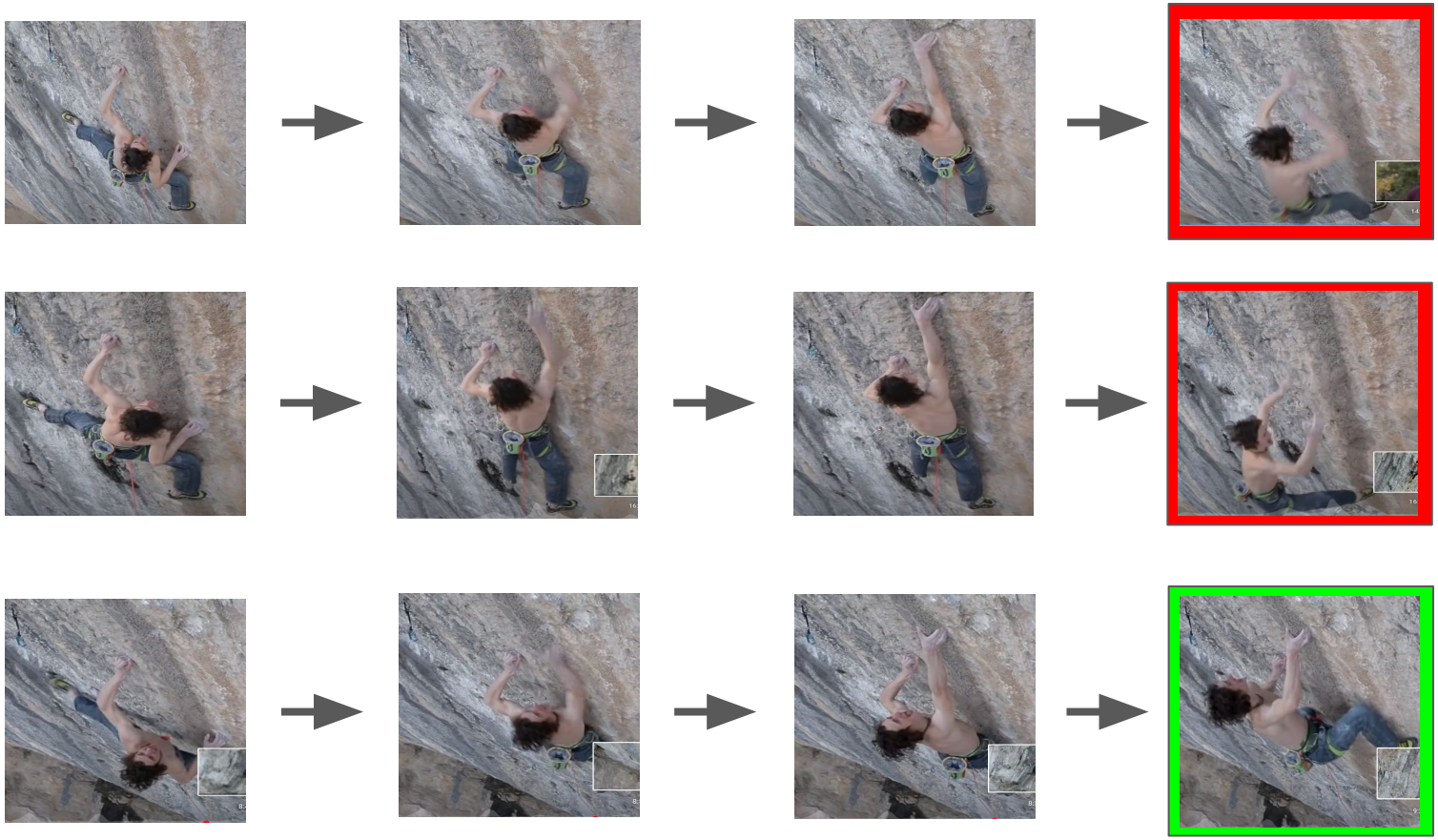
COME SI IMPARA?

ALGORITMO PER IL COMPUTER



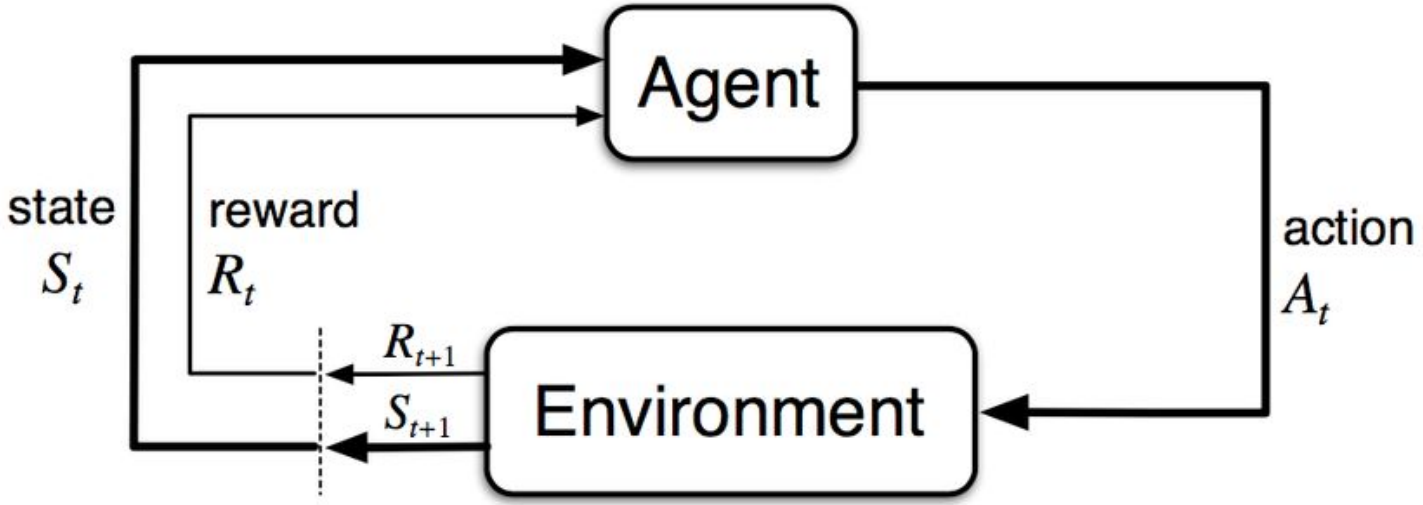
COME SI IMPARA?

PROVA ED ERRORE



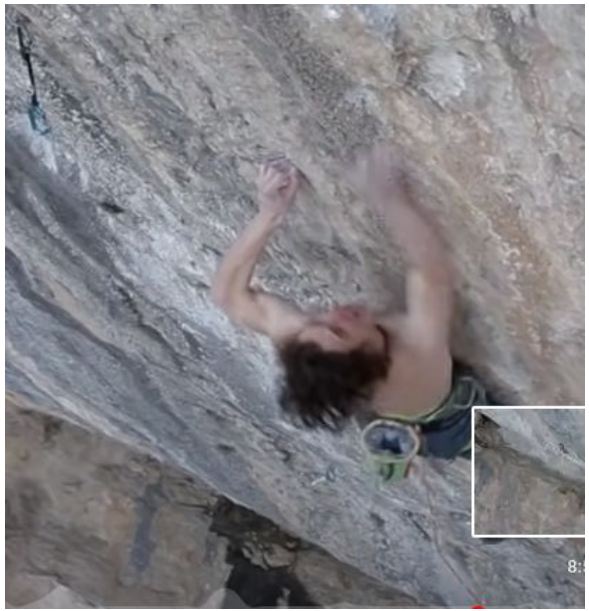
PROVA ED ERRORE

L'AMBIENTE DEL RL



LA DINAMICA DEL RL

STATE *t*



ACTION

ACTION

STATE *t+1*

REWARD = -1

STATE *t+1*

REWARD = +1

LA MATEMATICA DEL RL

$$q_*(s,a) = E[R_a + \gamma \max_{a'} q_*(s',a')]$$

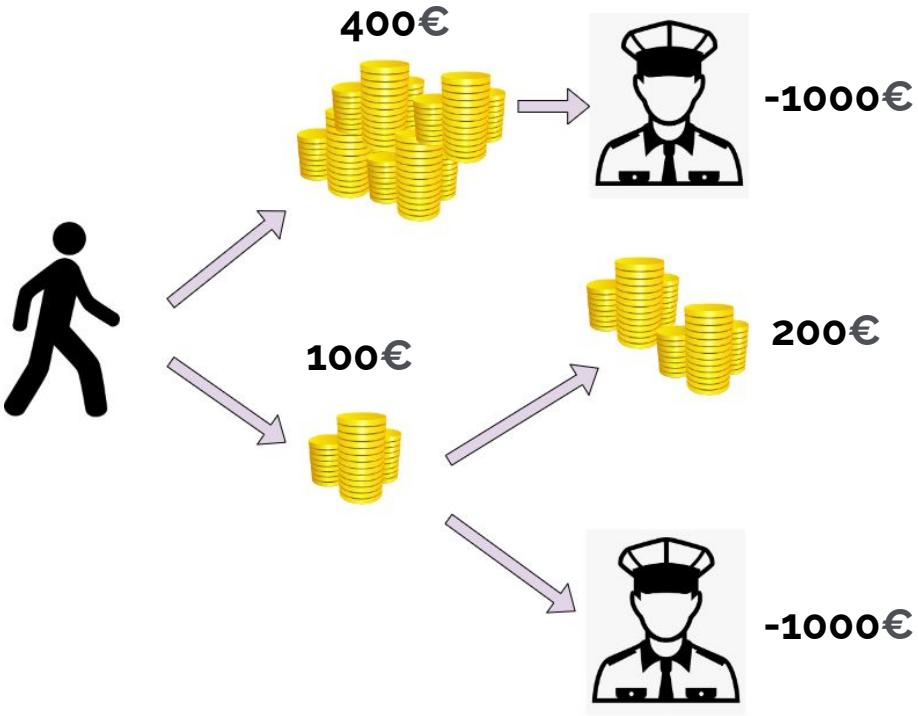
BELLMAN (1957)

LA MATEMATICA DEL RL

$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$

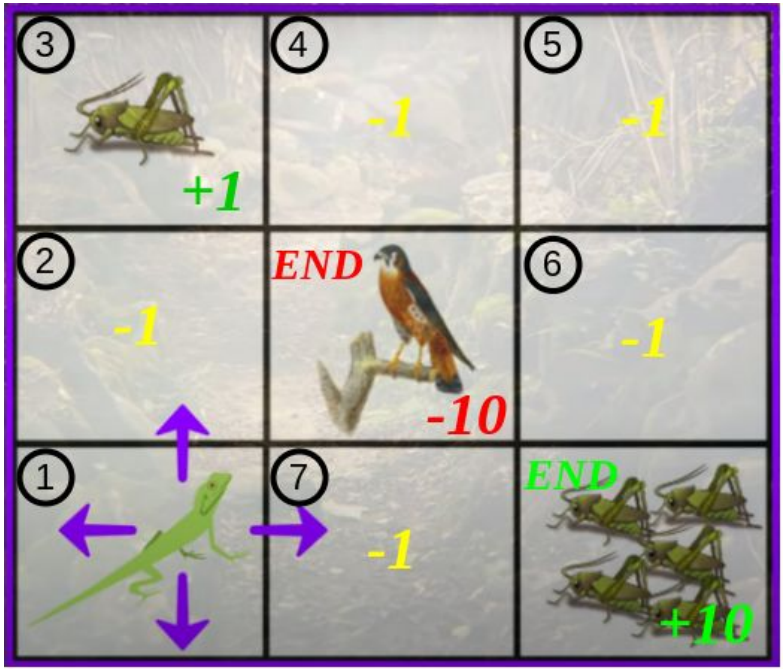
LA MATEMATICA DEL RL

$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$



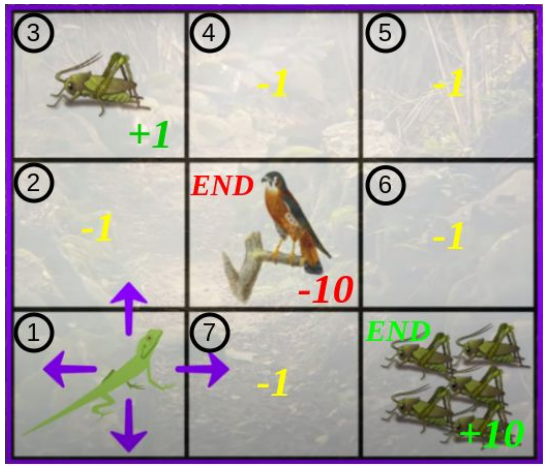
LA MATEMATICA DEL RL

$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$



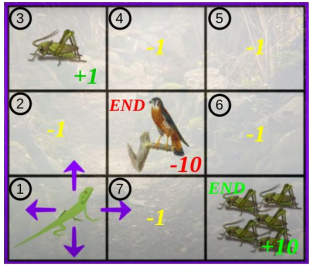
LA MATEMATICA DEL RL

$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$



	↑	↓	→	←
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0

LA MATEMATICA DEL RL



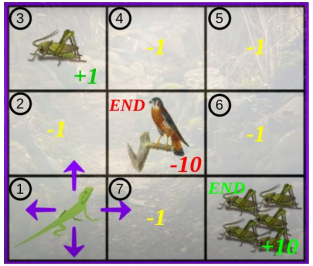
$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$

	↑	↓	→	←
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0

→

	↑	↓	→	←
1	-1	0	-1	0
2	+1	0	-10	0
3	0	-1	-1	0
4	0	-10	-1	+1
5	0	-1	0	-1
6	-1	+10	0	-10
7	-10	0	+10	0

LA MATEMATICA DEL RL



$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$

	↑	↓	→	←
1	-1	0	-1	0
2	+1	0	-10	0
3	0	-1	-1	0
4	0	-10	-1	+1
5	0	-1	0	-1
6	-1	+10	0	-10
7	-10	0	+10	0

→

	↑	↓	→	←
1	0	0	+9	0
2	+1	0	-10	0
3	0	0	0	0
4	0	-10	-1	+1
5	0	+9	0	0
6	-1	+10	0	-10
7	-10	0	+10	0

LA MATEMATICA DEL RL

	↑	↓	→	←
1	0	0	+9	0
2	+1	0	-10	0
3	0	0	0	0
4	0	-10	-1	+1
5	0	+9	0	0
6	-1	+10	0	-10
7	-10	0	+10	0

	↑	↓	→	←
1	0	0	+9	0
2	+1	+9	-10	0
3	0	0	0	0
4	0	-10	+8	+1
5	0	+9	0	0
6	+8	+10	0	-10
7	-10	0	+10	+9

	↑	↓	→	←
1	+8	0	+9	0
2	+1	+9	-10	0
3	0	+8	+7	0
4	0	-10	+8	+1
5	0	+9	0	+7
6	+8	+10	0	-10
7	-10	0	+10	+9

	↑	↓	→	←
1	+8	0	+9	0
2	+9	+9	-10	0
3	0	+8	+7	0
4	0	-10	+8	+9
5	0	+9	0	+7
6	+8	+10	0	-10
7	-10	0	+10	+9

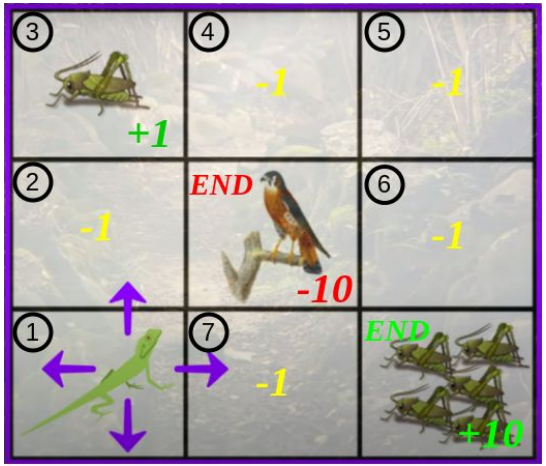
	↑	↓	→	←
1	+8	0	+9	0
2	+9	+9	-10	0
3	0	+8	+8	0
4	0	-10	+8	+9
5	0	+9	0	+8
6	+8	+10	0	-10
7	-10	0	+10	+9

=

	↑	↓	→	←
1	+8	0	+9	0
2	+9	+9	-10	0
3	0	+8	+8	0
4	0	-10	+8	+9
5	0	+9	0	+8
6	+8	+10	0	-10
7	-10	0	+10	+9

LA MATEMATICA DEL RL

$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$



...si può fare anche per tentativi...

	↑	↓	→	←
1	+8	0	+9	0
2	+9	+9	-10	0
3	0	+8	+8	0
4	0	-10	+8	+9
5	0	+9	0	+8
6	+8	+10	0	-10
7	-10	0	+10	+9

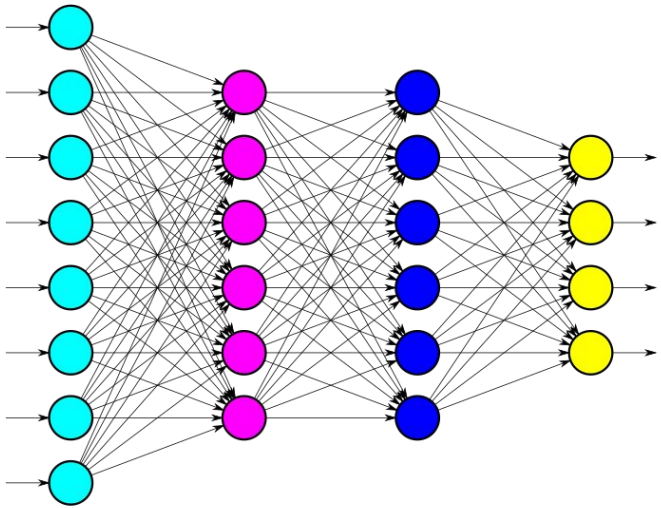
SEMPRE APPLICABILE?

$$\longrightarrow \quad |S| = 10^{123}$$

$$\longrightarrow \quad |S| = ??$$

MACHINE LEARNING

$$q_*(s,a) = R_a + \max_{a'} q_*(s',a')$$



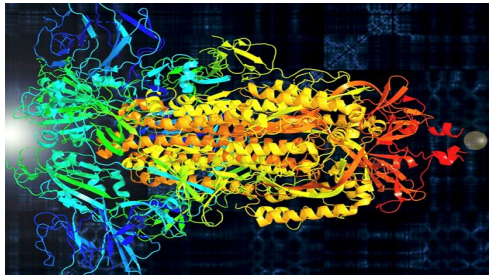
LA MATEMATICA DEL RL

$$q_*(s,a) = E[R_a + \gamma \max_{a'} q_*(s',a')]$$

DEEPMIND

ALPHA GO (2016)

ALPHA STAR (2019)



ALPHA FOLD (2021)

Grazie dell'attenzione!

E-MAIL davide.basso@phd.units.it

UNIVERSITÀ DEGLI STUDI DI TRIESTE