

Deep Q Learning: From Paper to Code

What We Have Learned

Phew ...



Agents, Actions, Environments



Markov Decision Processes



$$(S_1, A_1, R_1, S_2, A_2, R_2, \dots)$$

State depends only on
previous state and action

Markov Decision Process

Optimal Policies

$$v_{\pi}(s) = \sum_a \pi(a, s) \sum_{s', r} p(s', r | s, a) [r + \gamma v_{\pi}(s')]$$

Bellman Equation

Recursive relationship between value functions

$$v_*(s) = \max_a \sum_{s', r} p(s', r | s, a) [r + \gamma v_*(s')]$$

$$q_*(s, a) = \sum_{s', r} p(s', r | s, a) [r + \gamma \max_{a'} q_*(s', a')]$$

Q Learning

Model Free → Q Learning



Estimate p by playing

Off policy temporal difference learning method

Tabular Learning



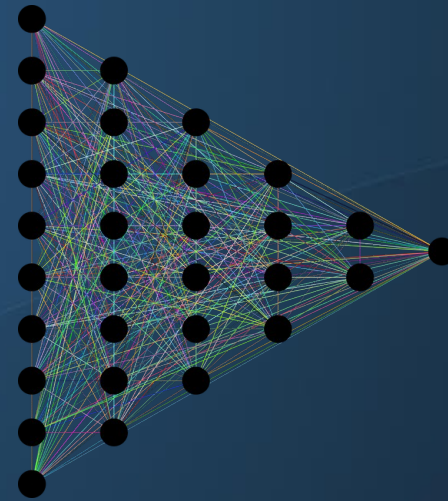
Small number of states, actions

	Action 1	Action 2	Action n
State 1	$Q(1,1)$	$Q(1,2)$	$Q(1,n)$
State 2	$Q(2,1)$	$Q(2,2)$	$Q(2,n)$
State m	$Q(m,1)$	$Q(m,2)$	$Q(m,n)$

Deep Learning For Large State Spaces



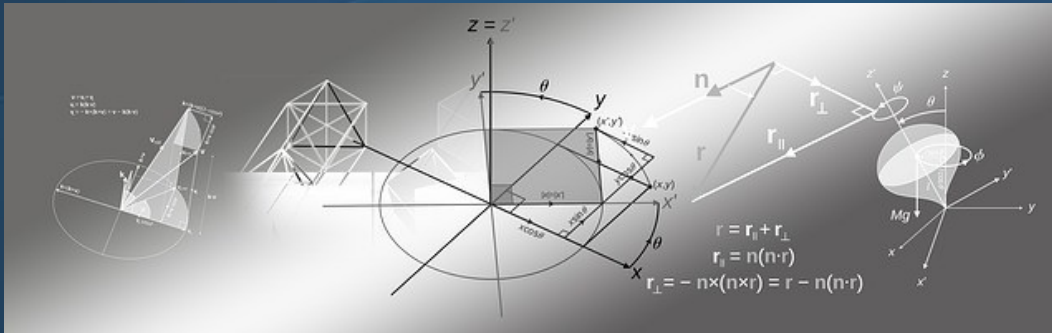
Tables will not work



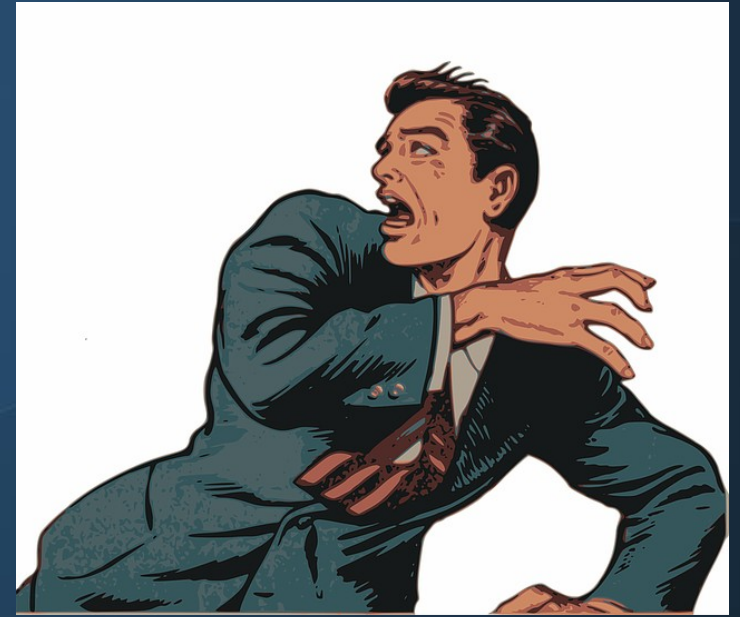
Deep neural networks

Simple DNN implementation failed!

Need For New Algorithm



Correlations ruin learning



Chasing a moving target

Code Clean Up



Argparse and inheritance

Up Next

