Playing Snake with Q Learning

Paul Mora Sancho

DATA SCIENCE RETREAT

Agenda

1. Q Learning Introduction
2. Q Tables
3. Deep Learning
4. Human vs. Machine

Agenda

1. Q Learning Introduction
2. Q Tables
3. Deep Learning
4. Human vs. Machine

Let us consider an example

States

Dressed Up as…

Observation

Actions

Going to…

Observation

How do we know what to do when dressed in a certain way?

The Oracle predicts our reward for an action

State

Action

Predict

Q Values

Medium

High

Low

Low Medium

Agenda

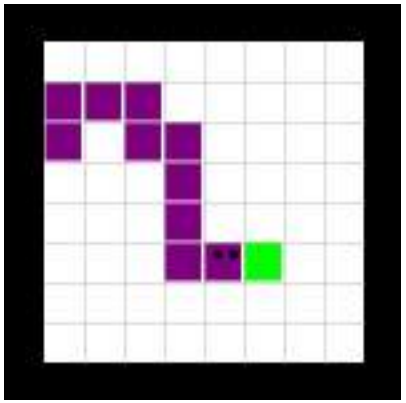
1. Q Learning Introduction
2. **Q Tables**
3. Deep Learning
4. Human vs. Machine

Q Tables

General Information

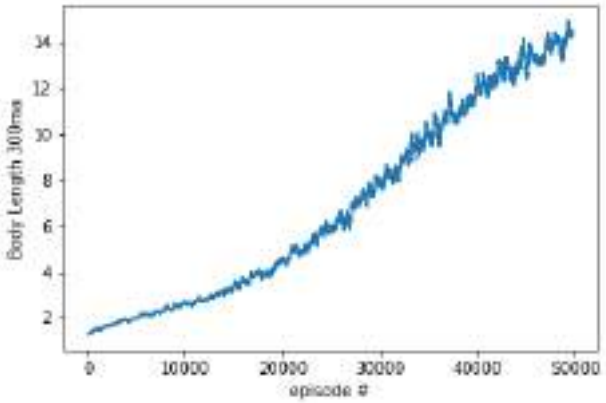
- **Value based** algorithm in reinforcement learning
- Fancy word for a **lookup table**
- Calculating **maximum future reward**
- **Iterative process**, as Q-Table needs improvement
- Makes use of **Bellman equation**
- **Initial values** of Q-Tables **are zero**, but there are **multiple** approaches → Optimal intrinsic values

Performance Preview



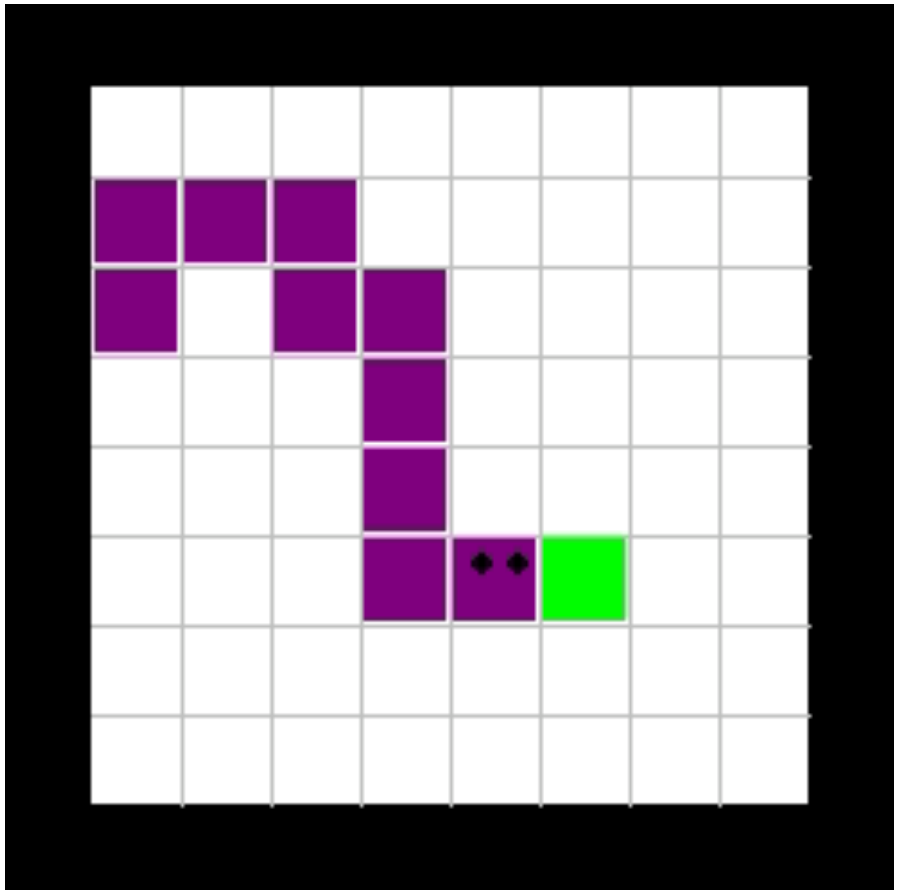
Workings overview

Reward Function



Q Tables - Example

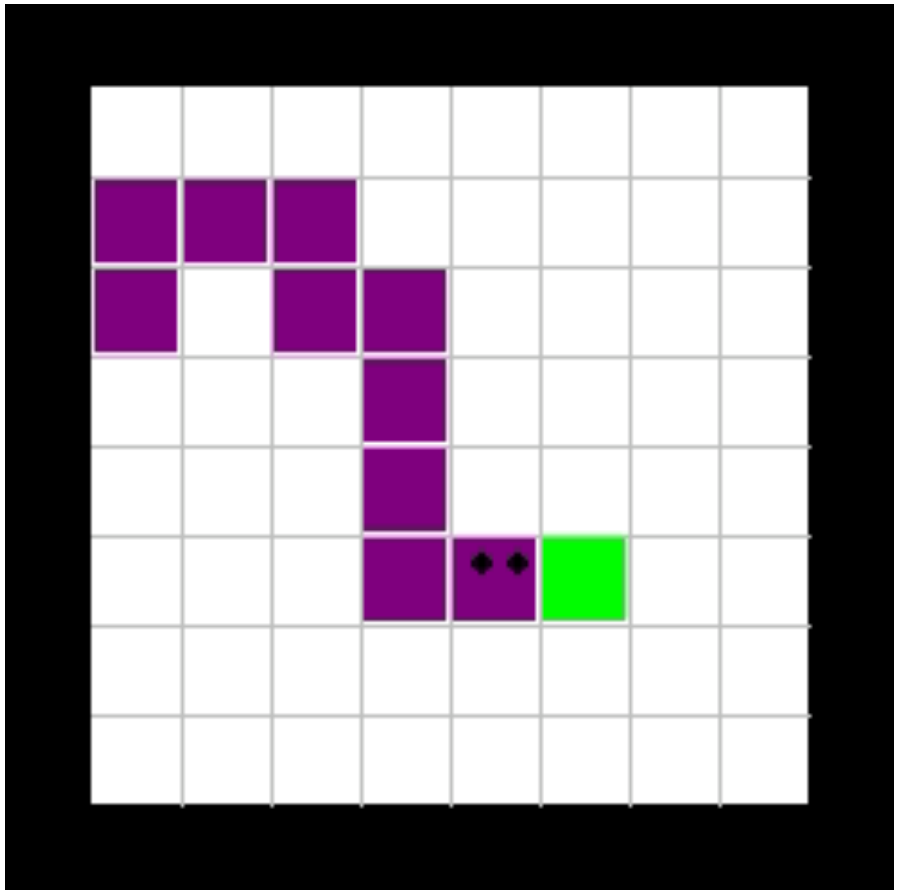
Situation



Procedure

Q Tables - Example

Situation



Procedure

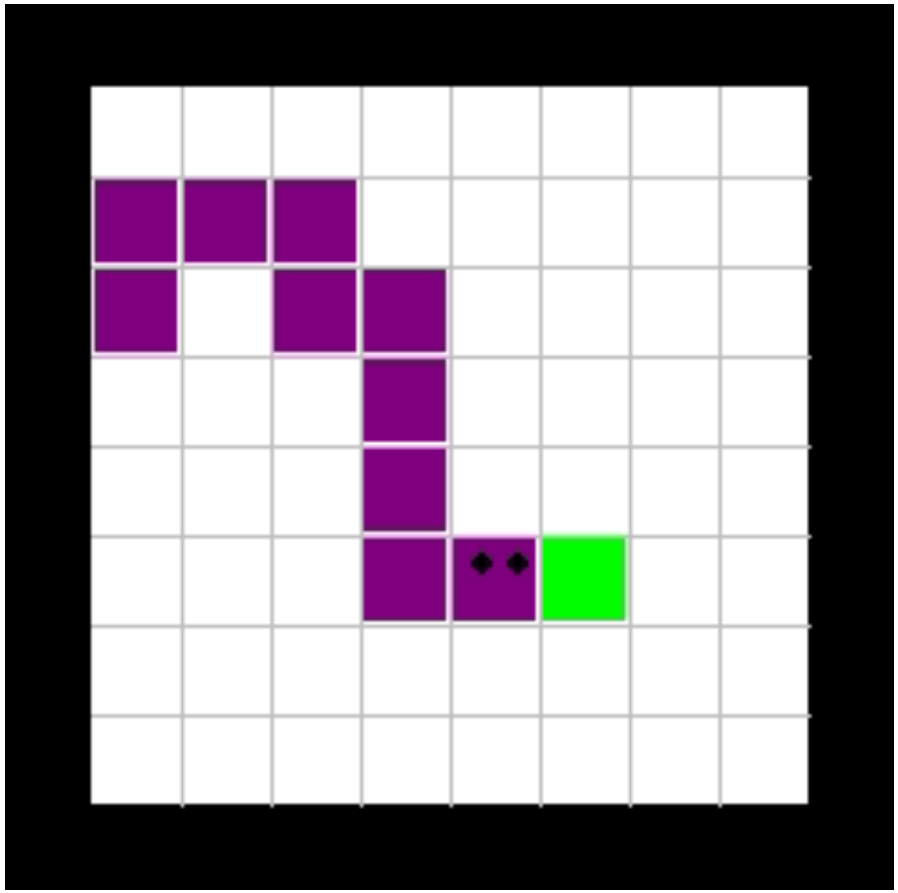
Danger Variables				Snack Variables	
Up	Down	Left	Right	Angle	Distance
0	0	1	0	90	1

Try different actions given this state

$$Q(s,a) = r + \gamma(max(Q(s',a'))$$

Q Tables - Example

Situation



Procedure

Danger Variables				Snack Variables	
Up	Down	Left	Right	Angle	Distance
0	0	1	0	90	1

↓ Try different actions given this state

$$Q(s,a) = r + \gamma(max(Q(s',a'))$$

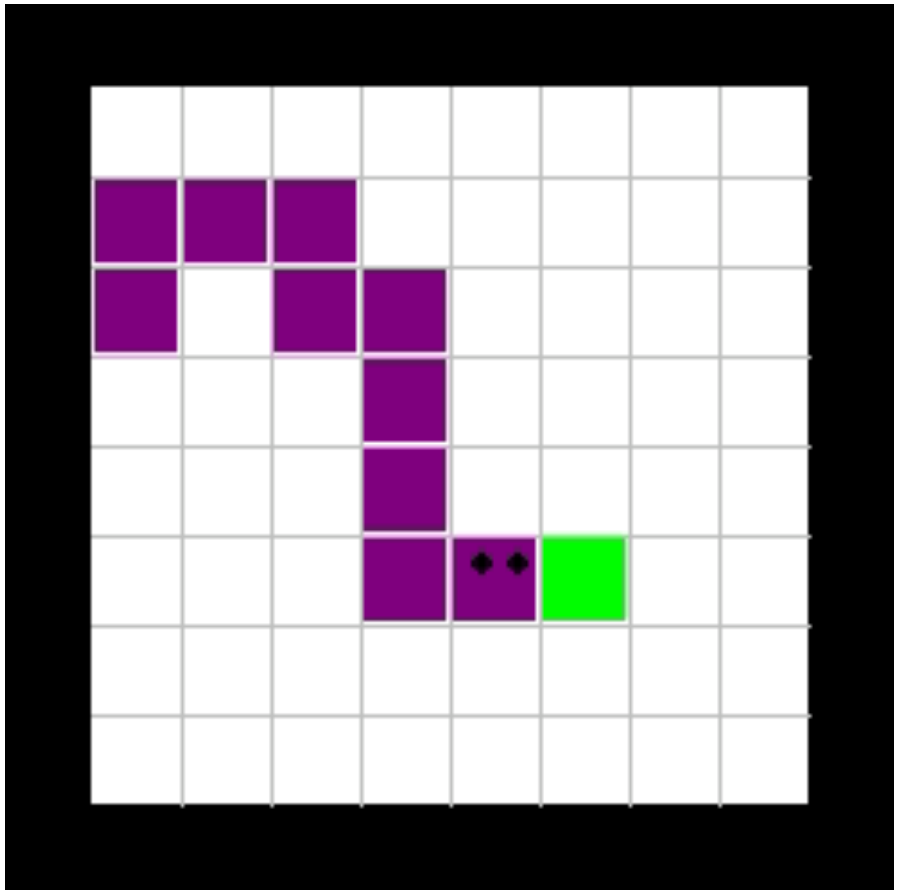
↓ Calculate Q Values

Q Values

Up	Down	Left	Right
0.86201	0.87409	-1	1

Q Tables - Example

Situation



Procedure

Danger Variables				Snack Variables	
Up	Down	Left	Right	Angle	Distance
0	0	1	0	90	1

↓ Try different actions given this state

$$Q(s,a) = r + \gamma(max(Q(s',a'))$$

↓ Calculate Q Values

Q Values

Up	Down	Left	Right
0.86201	0.87409	-1	1

↓ Taking the highest Q Value

Right
1

Agenda

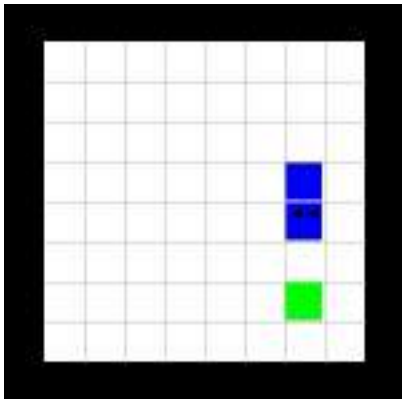
1. Q Learning Introduction
2. Q Tables
3. **Deep Learning**
4. Human vs. Machine

Deep Learning

General Information

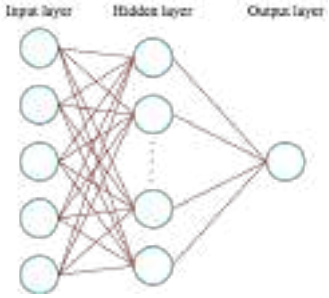
- **Model based** approach
- Goal is that algorithm **predicts** whether a movement in a certain direction gets the snake **closer to the apple**
- Data is gathered by **purely random movements** in the beginning and then **trains in epochs**
- **Trade-off** between too many variables and state information

Performance Preview



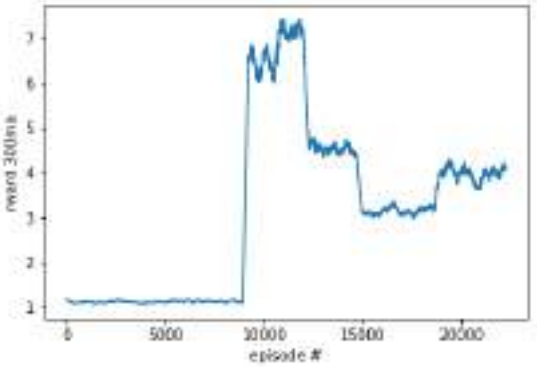
Input & Output Variables

- Danger proximity variables
- Apple location information
- Movement



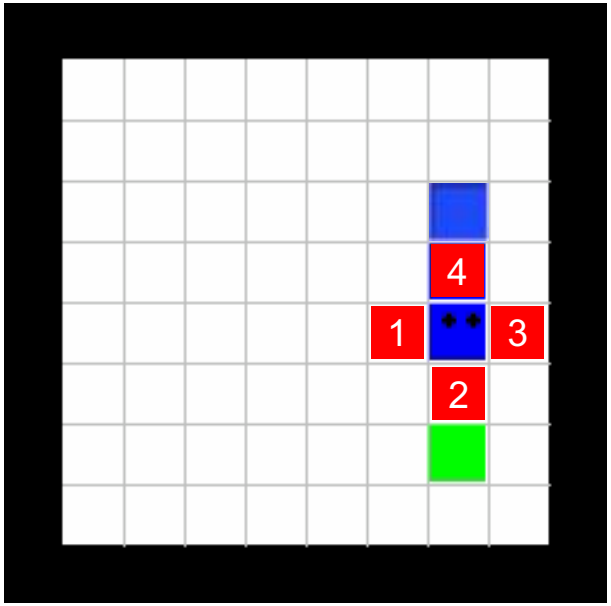
- 1: Getting closer to the apple
- 0: Getting further from the apple
- -1: Dying

Reward Function



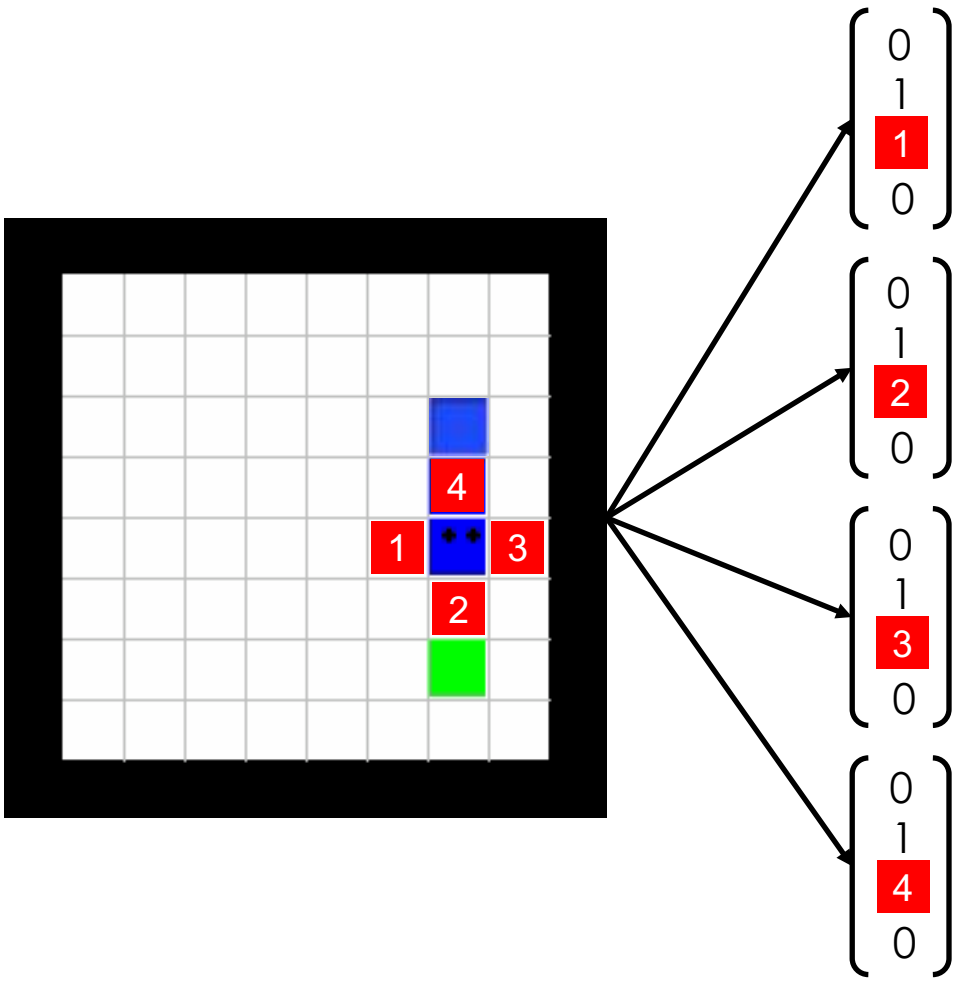
Q Tables - Example

Situation	Input of DL Model	Output	Action



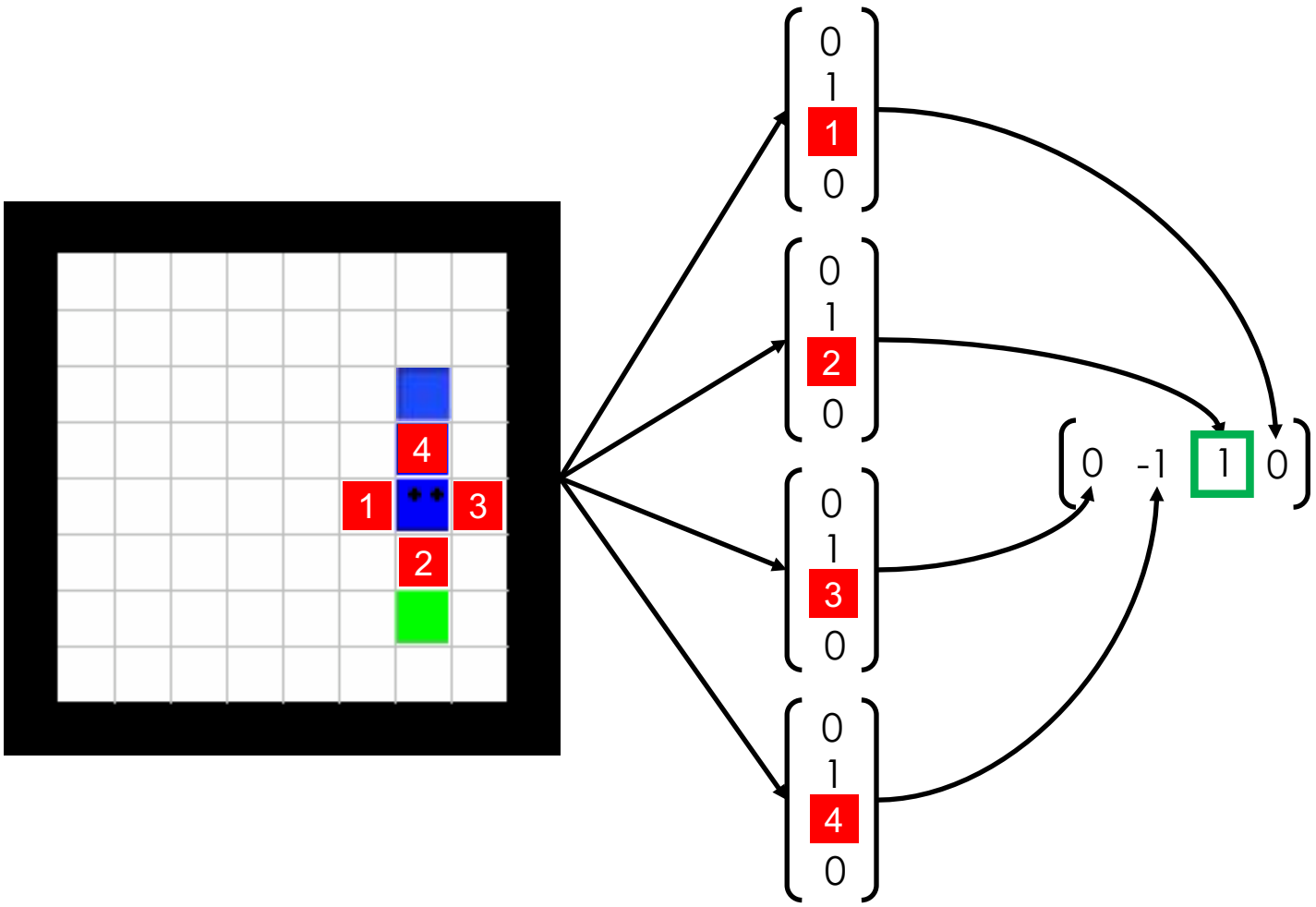
Q Tables - Example

Situation	Input of DL Model	Output	Action



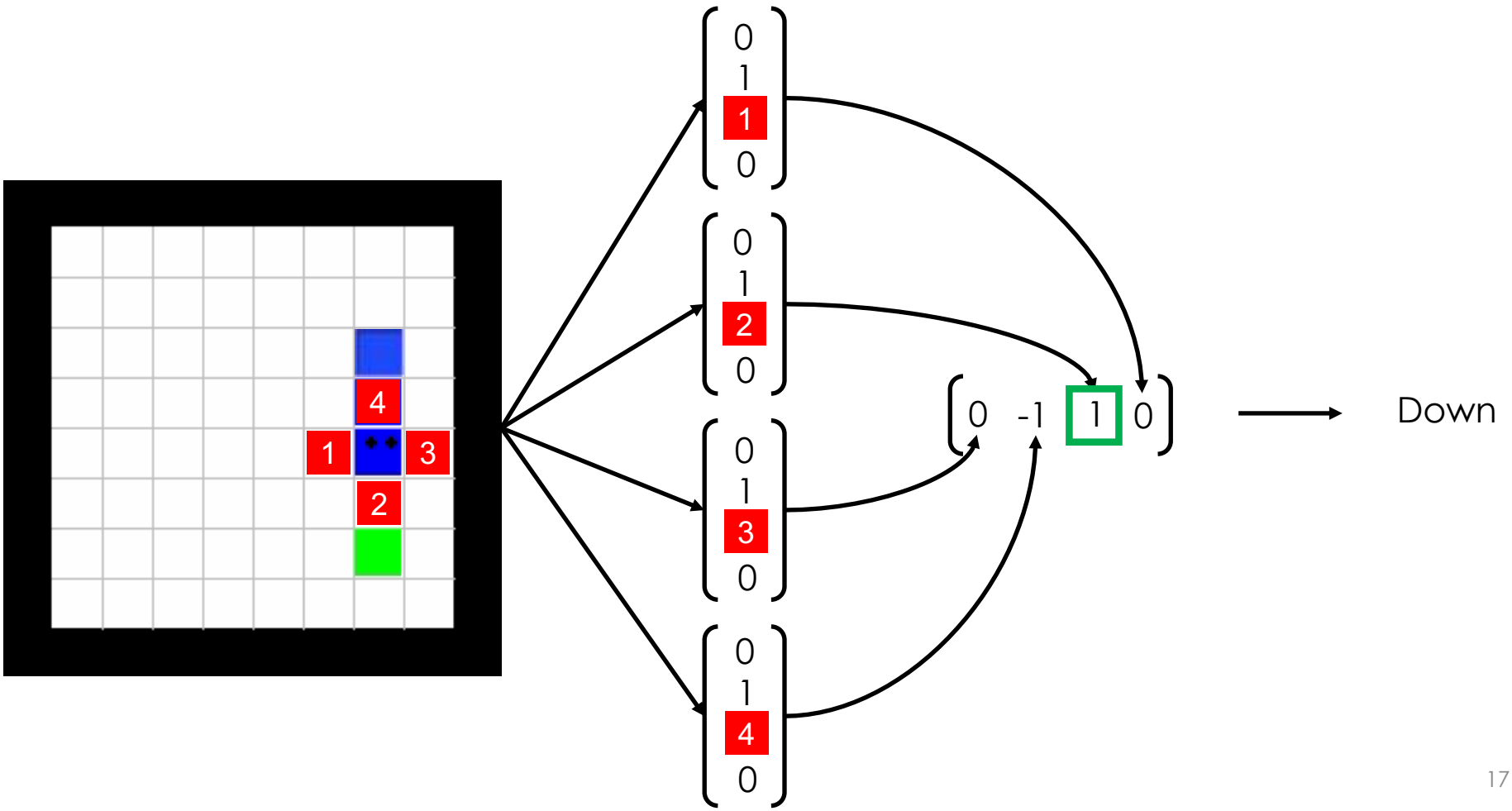
Q Tables - Example

Situation | Input of DL Model | Output | Action



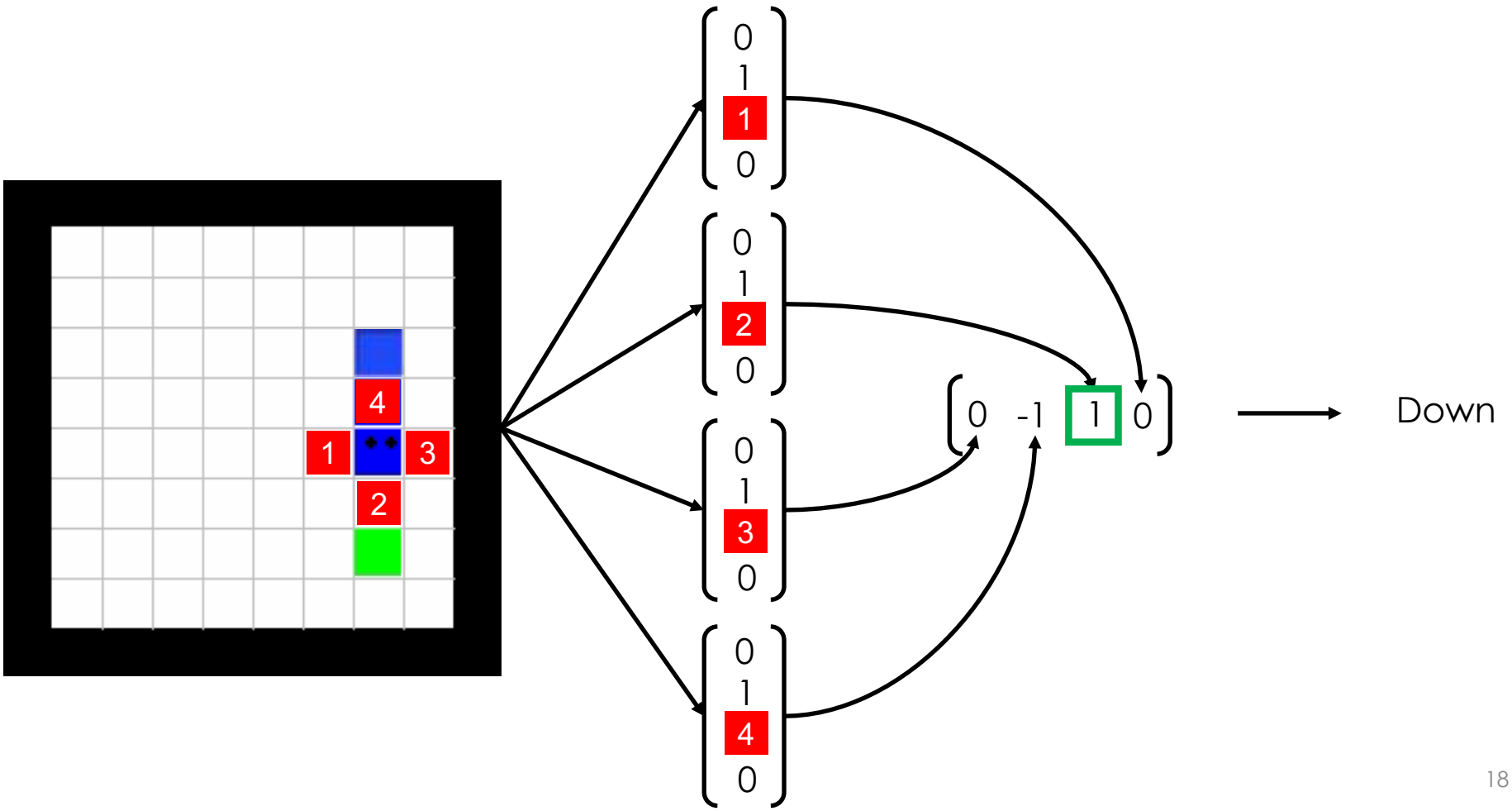
Q Tables - Example

Situation | Input of DL Model | Output | Action



Q Tables - Example

Situation | Input of DL Model | Output | Action

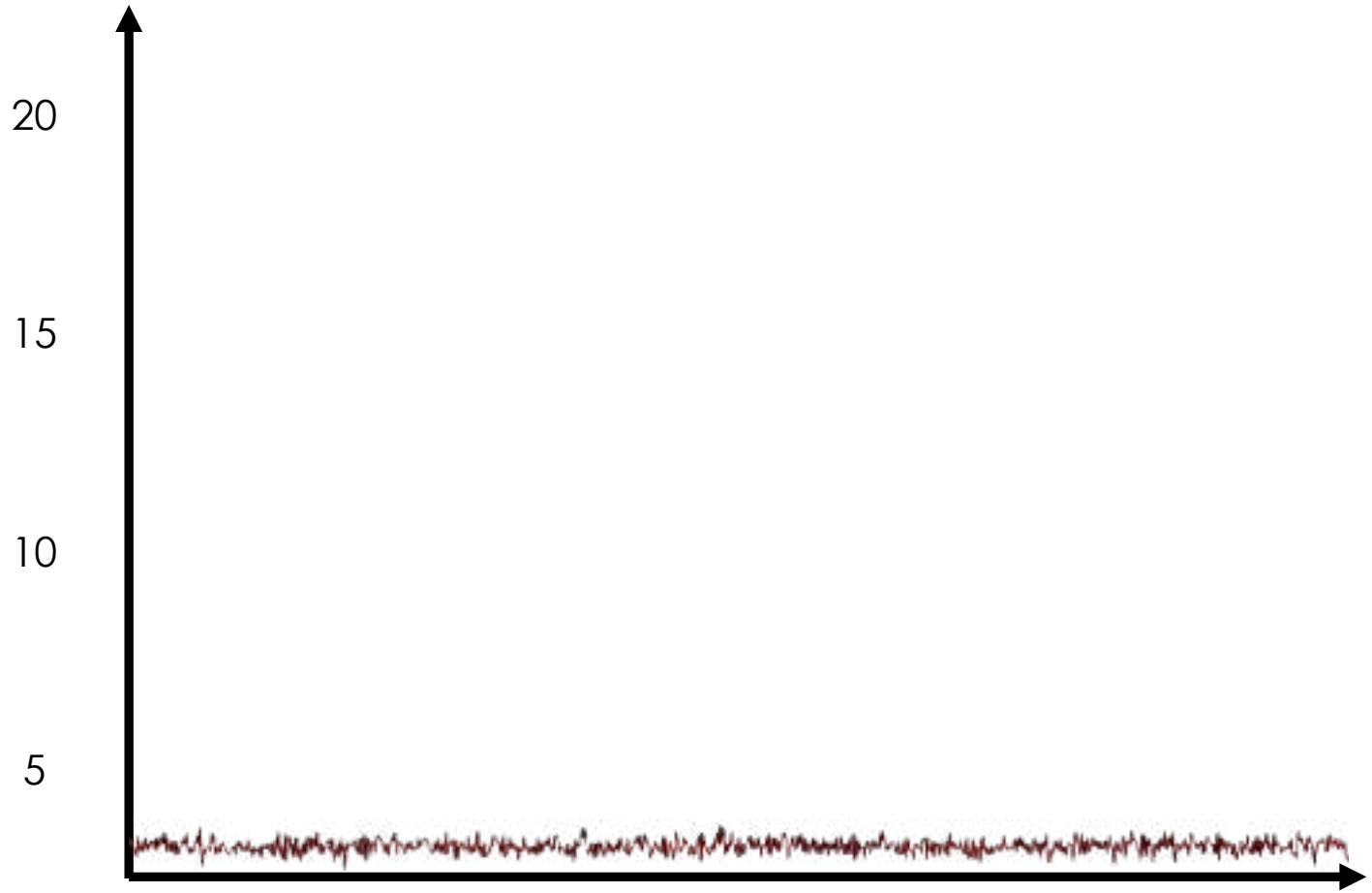


Agenda

1. Q Learning Introduction
2. Q Tables
3. Deep Learning
4. **Human vs. Machine**

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

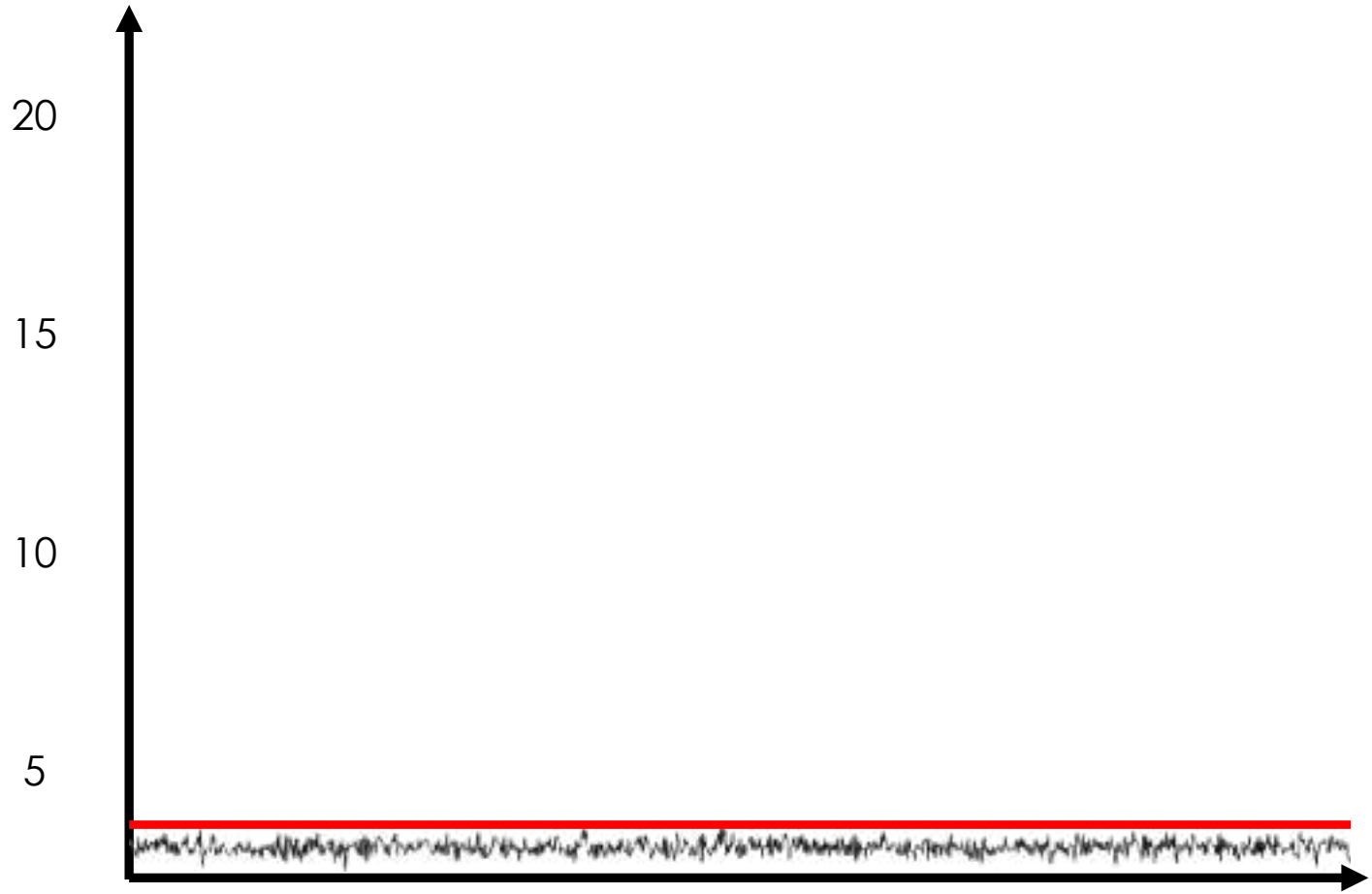
Random Policy

Score

1.5

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

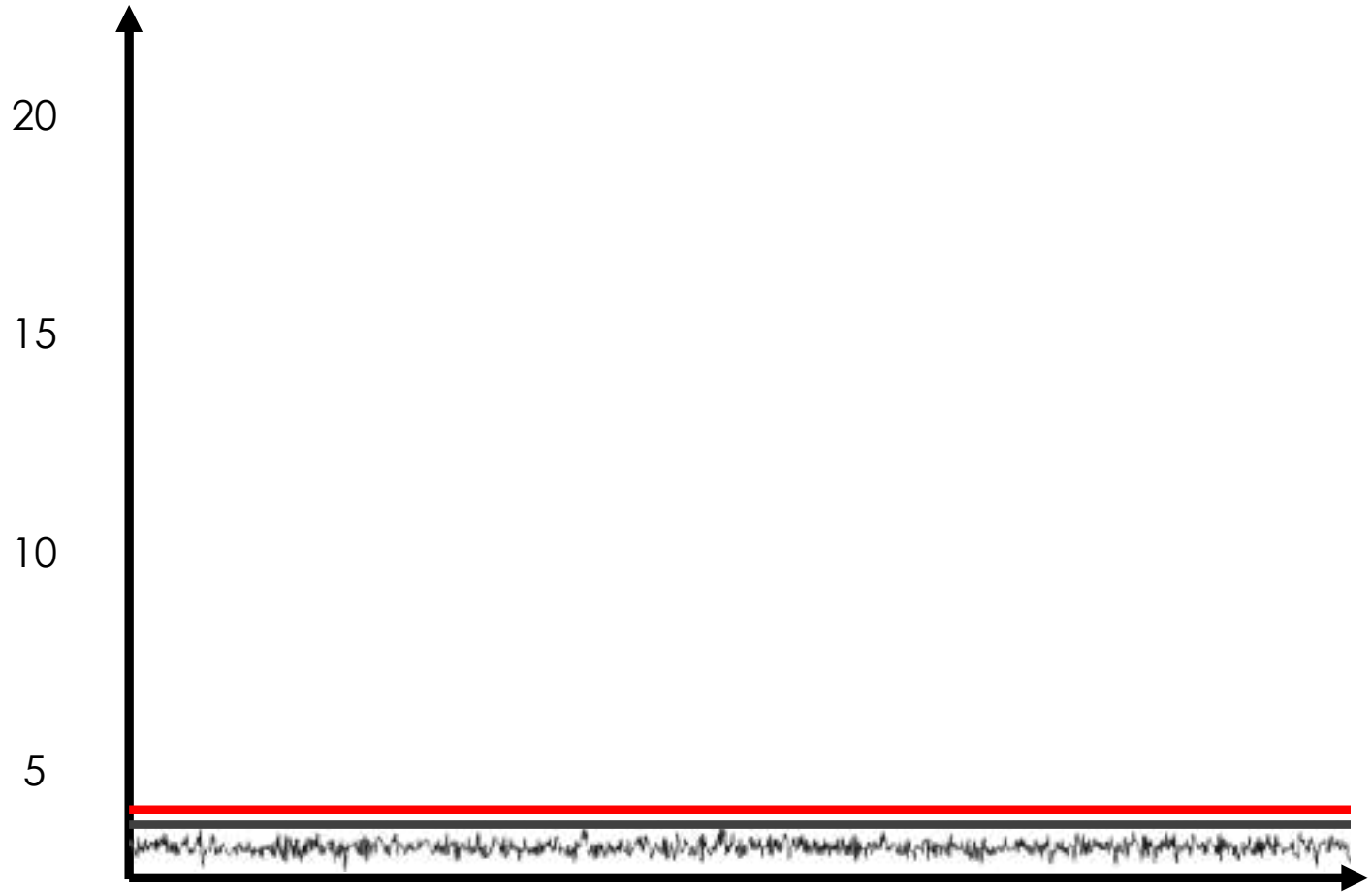
Samson

Score

3.5

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

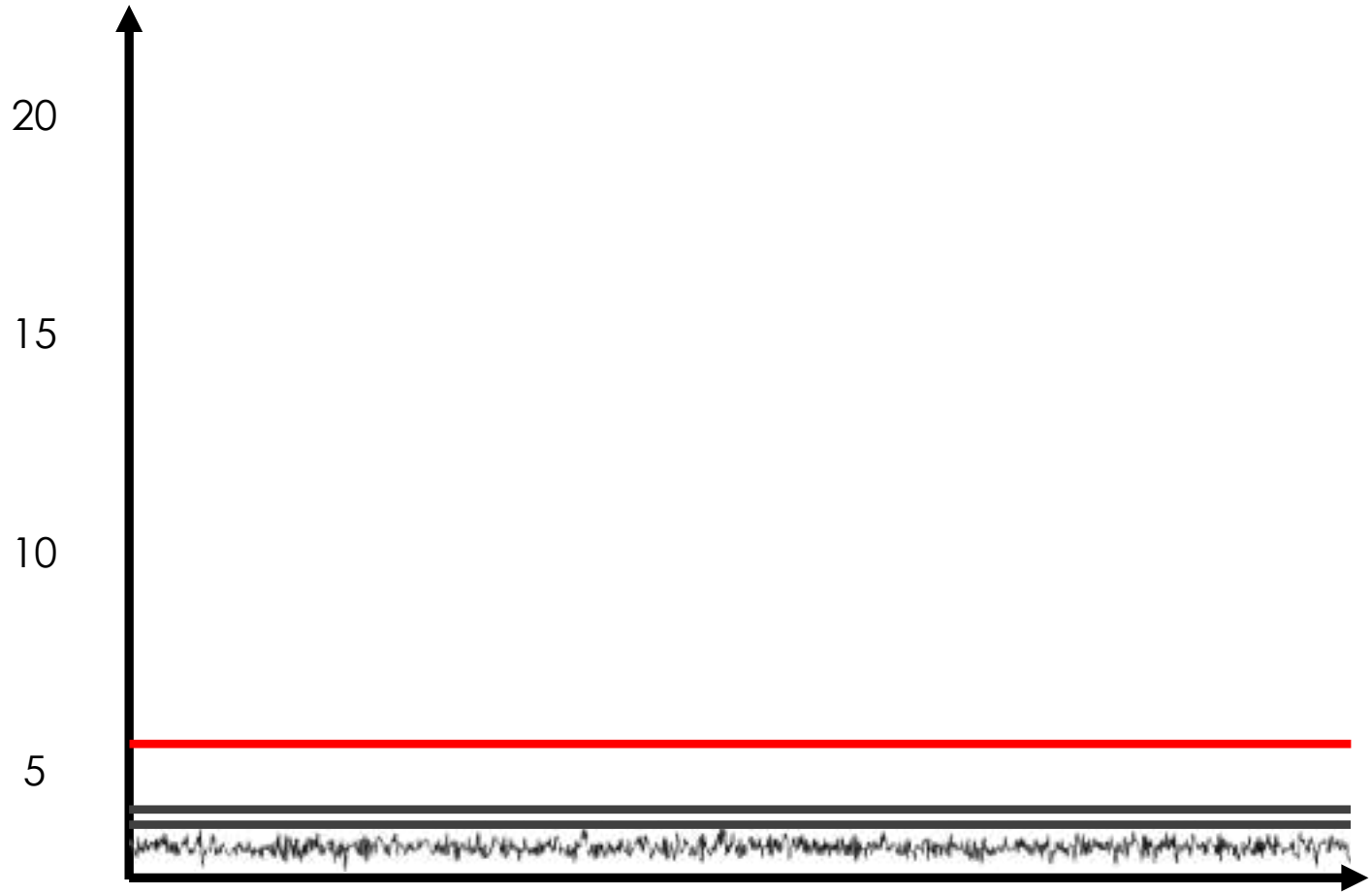
Aarun

Score

3.7

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

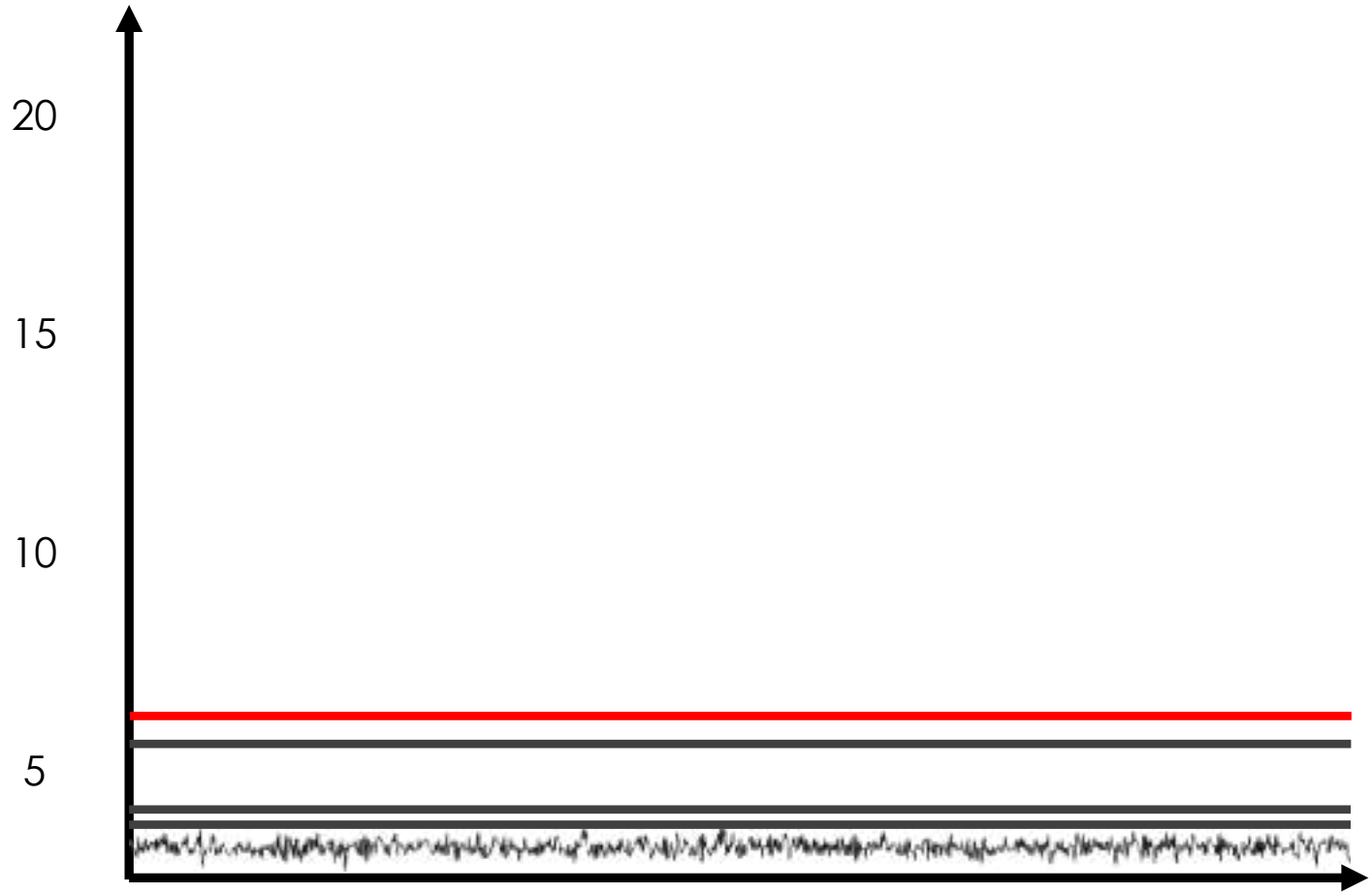
Rachel

Score

5.1

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

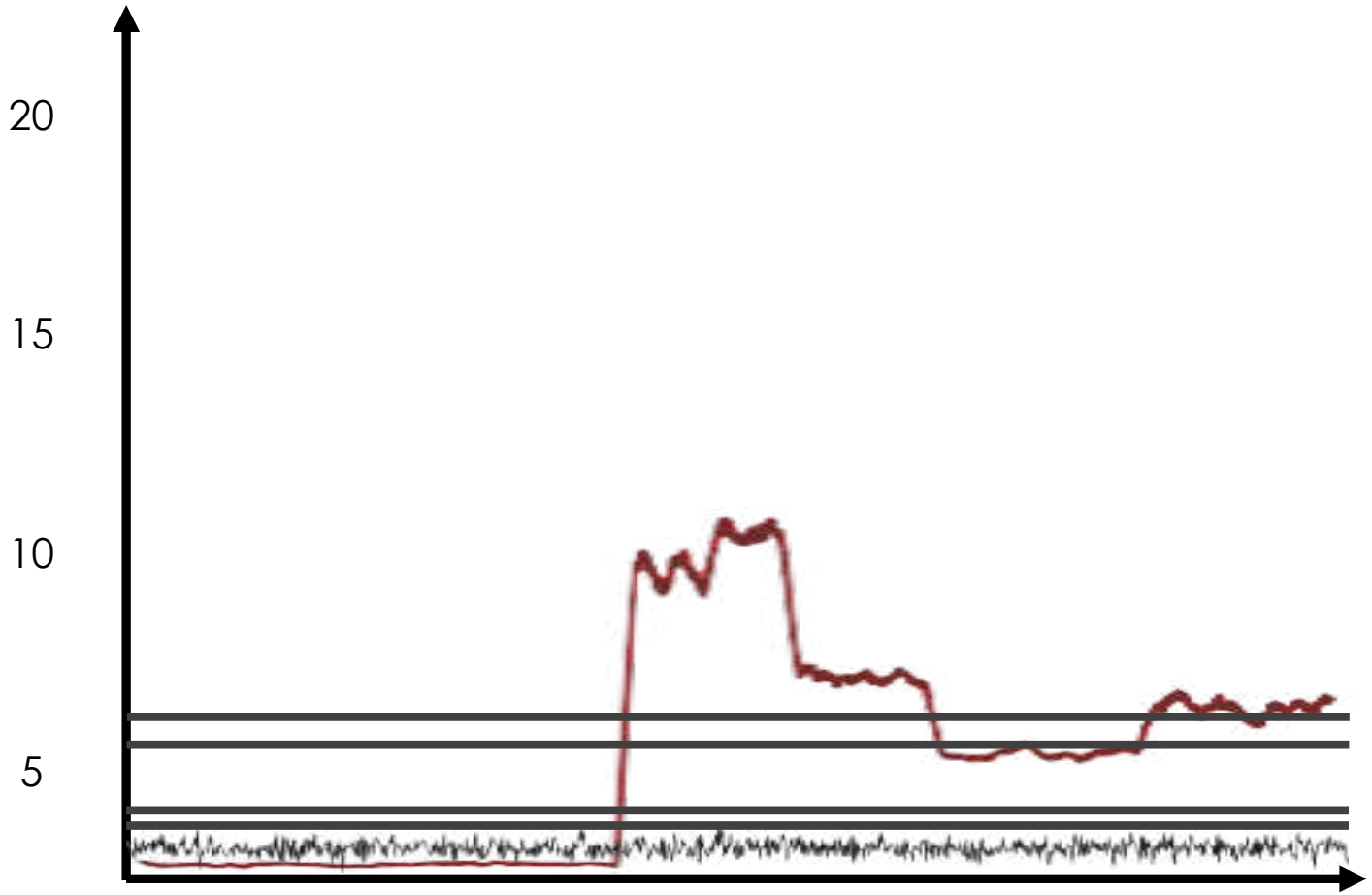
Dino

Score

6.9

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

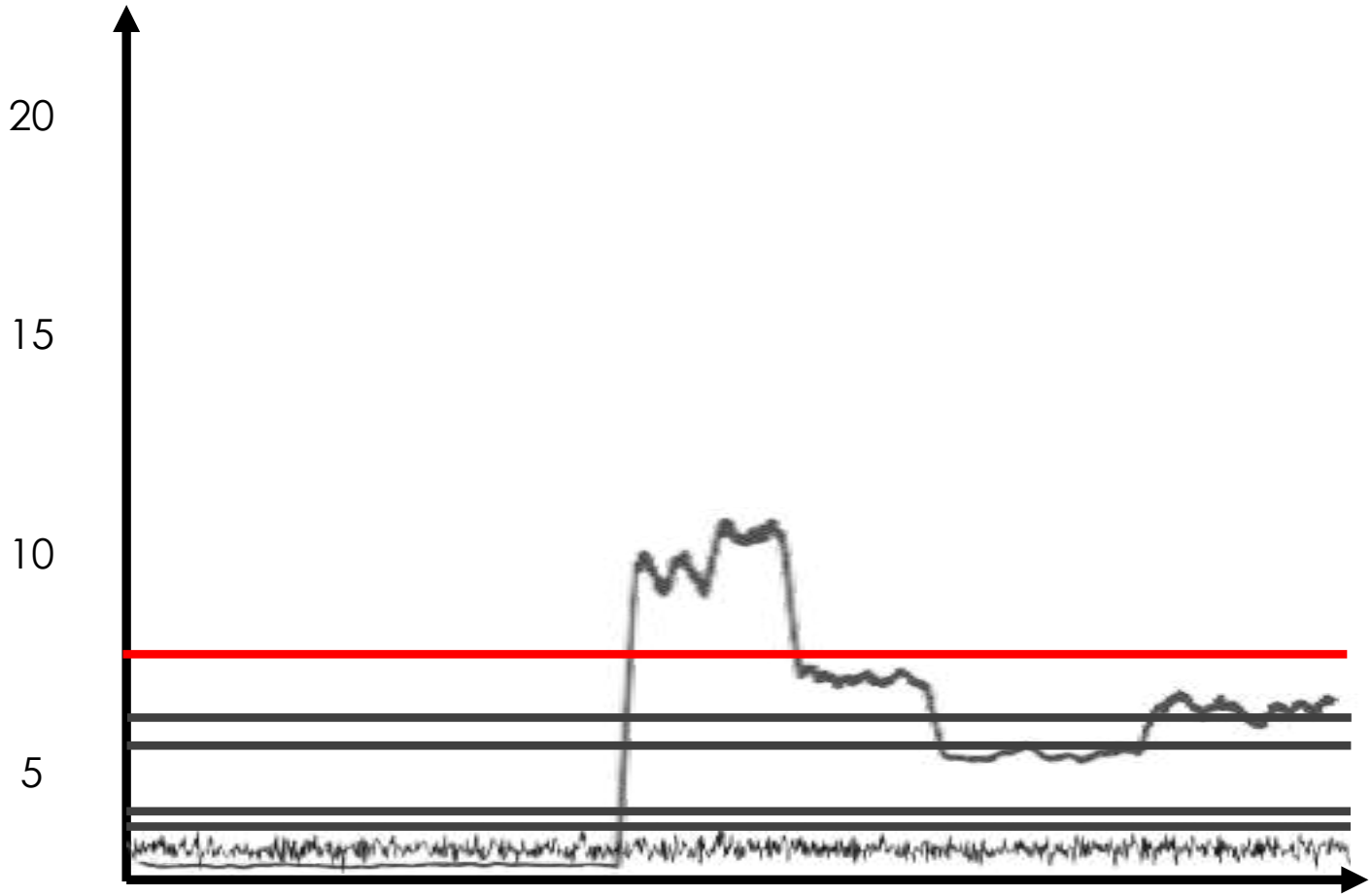
Deep Learning
(100k)

Score

7.2

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

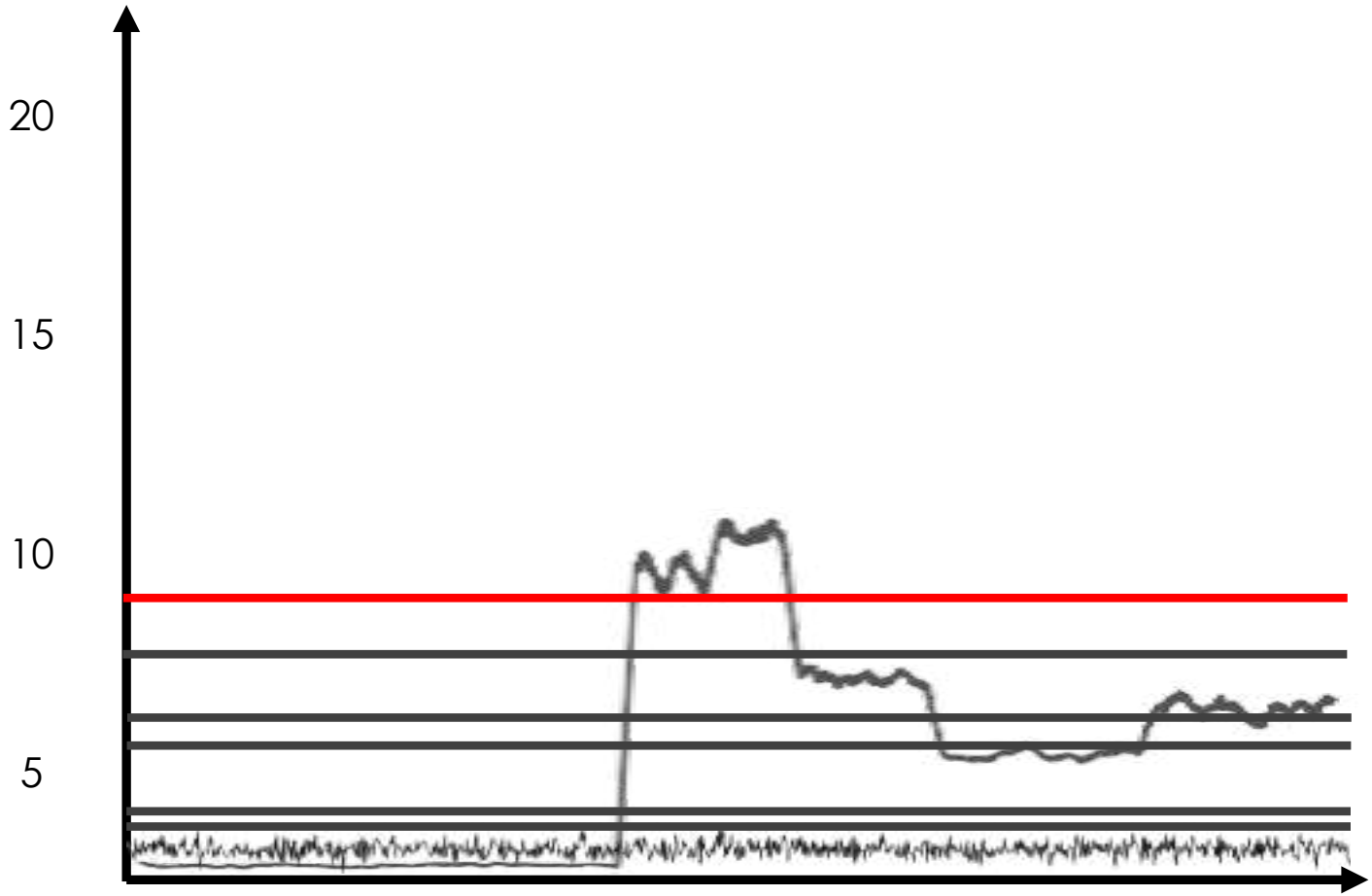
Stathis

Score

8.2

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

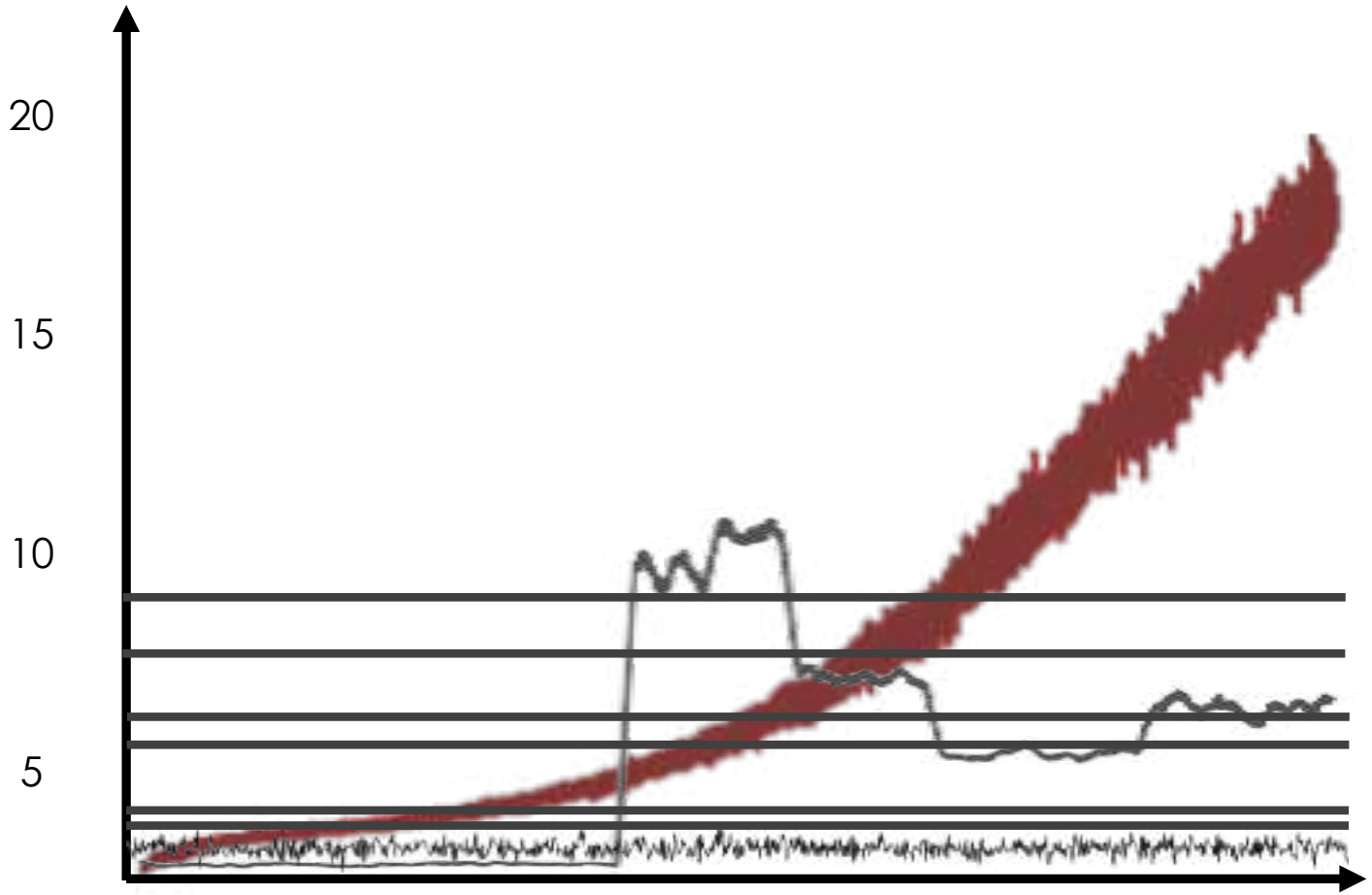
Peter

Score

8.7

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

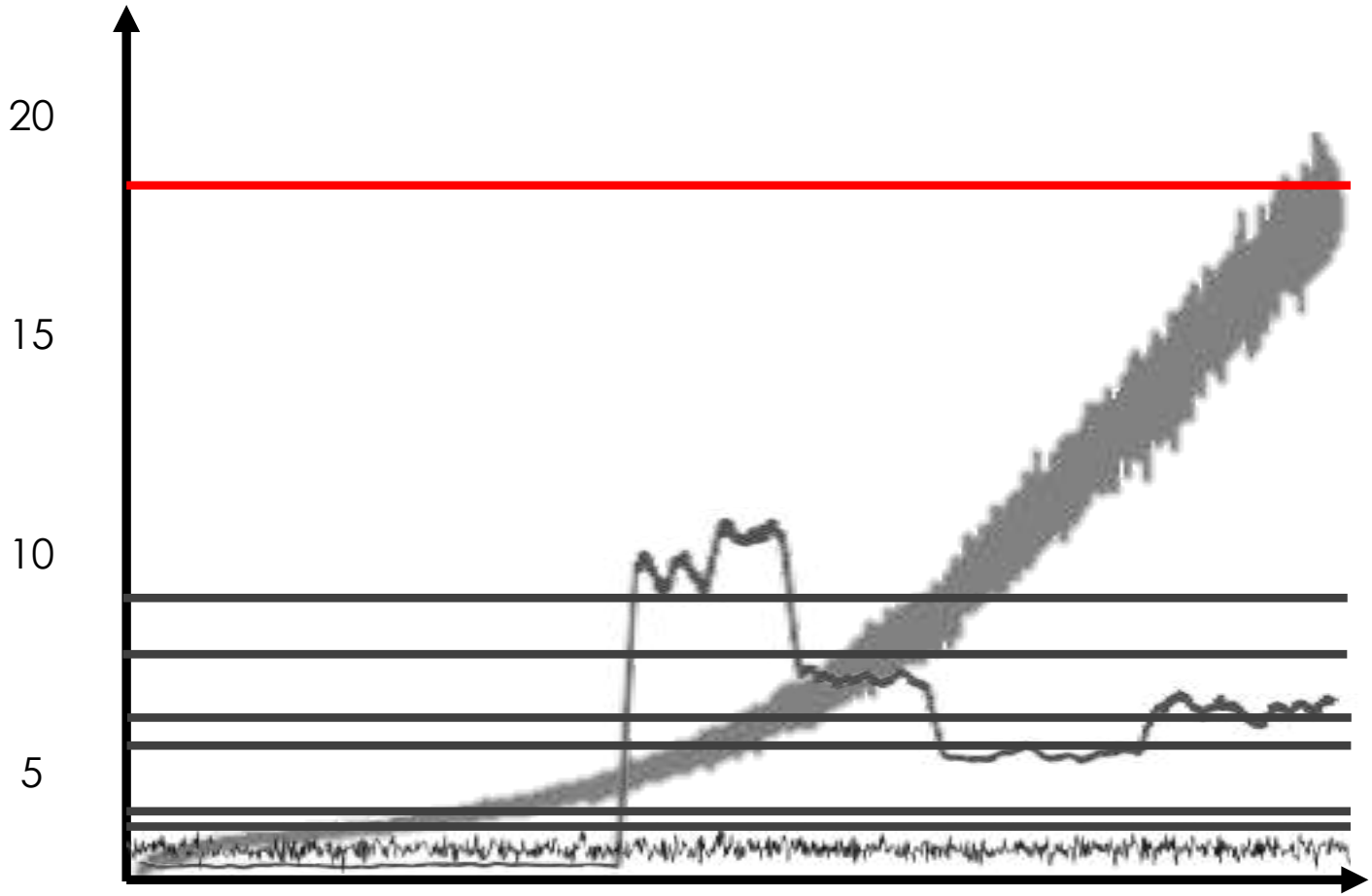
Q Tables
(200k)

Score

17.6

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

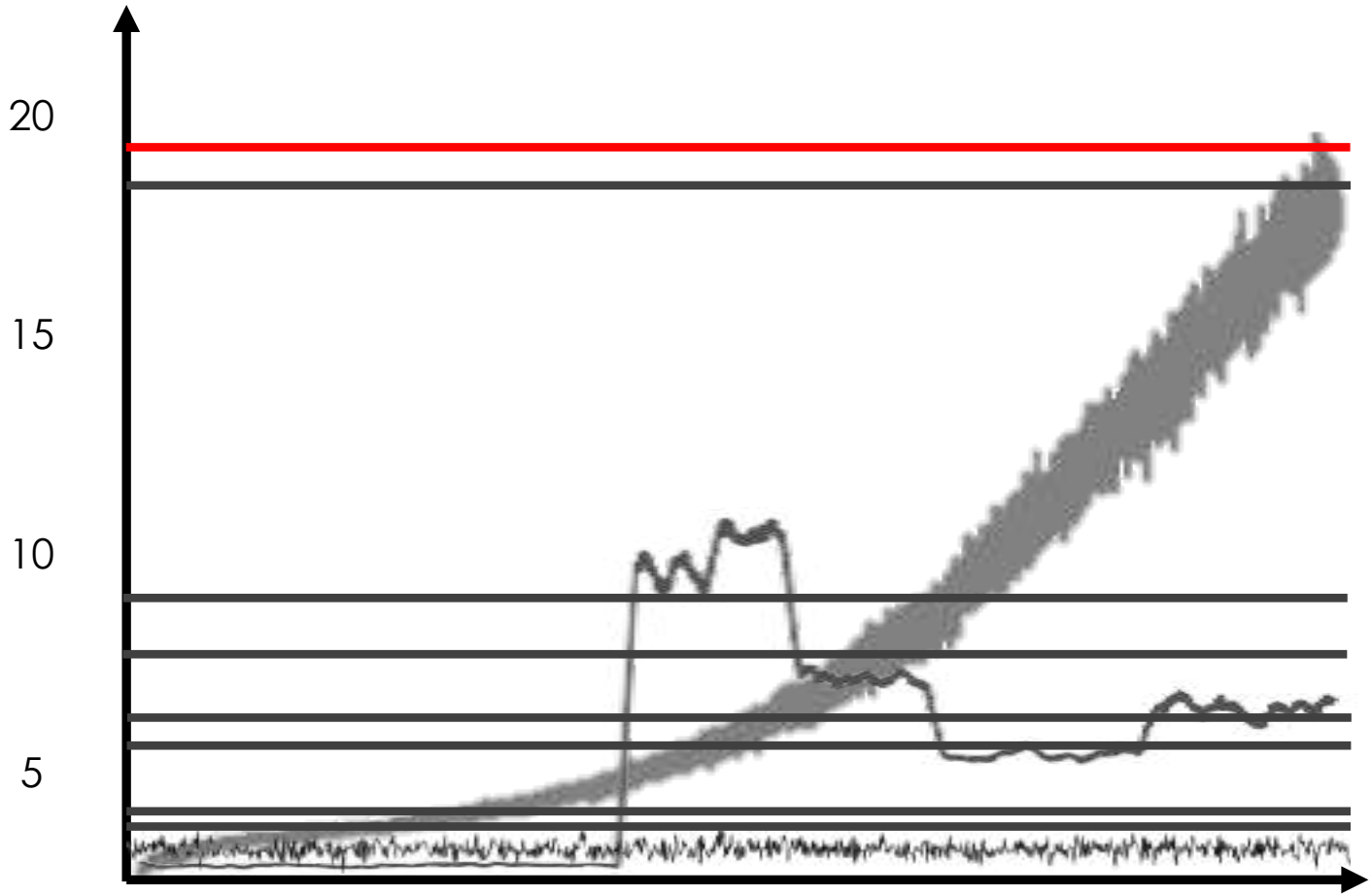
Stas

Score

18.2

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

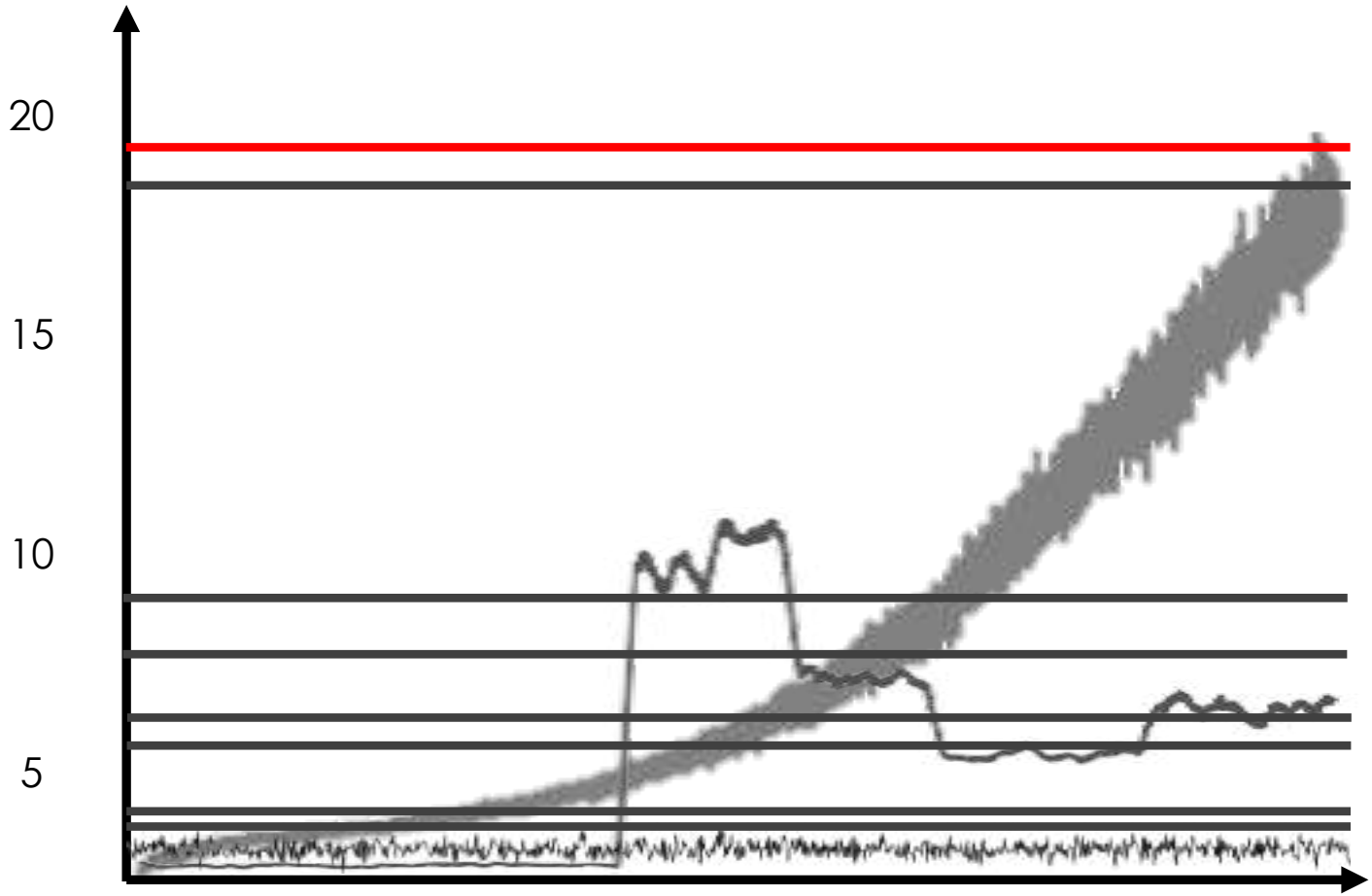
Adam

Score

19.2

The results of the fight human vs. machine

Graphical Representation



Algorithm/ Person

Adam

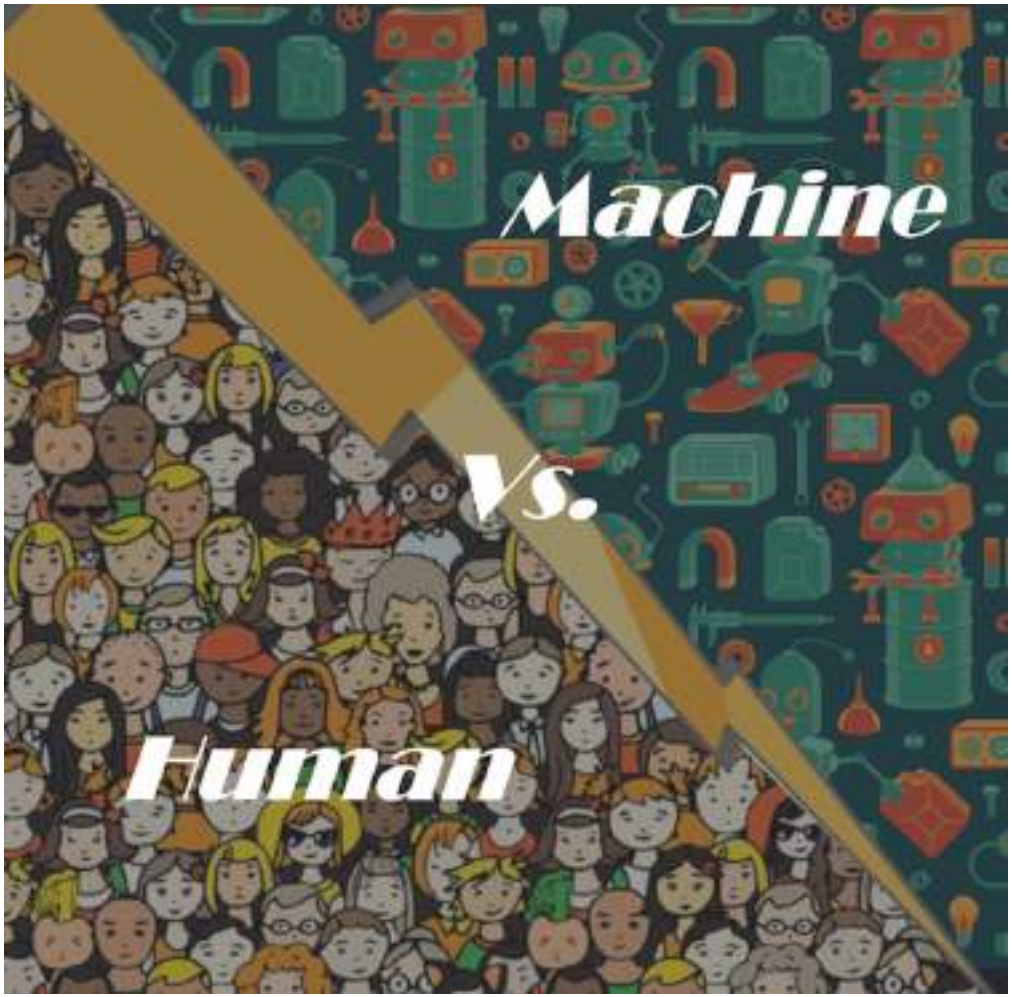
Score

19.2

Thank You

Email: paulmora@t-online.de
LinkedIn: Paul Michael Mora Sancho

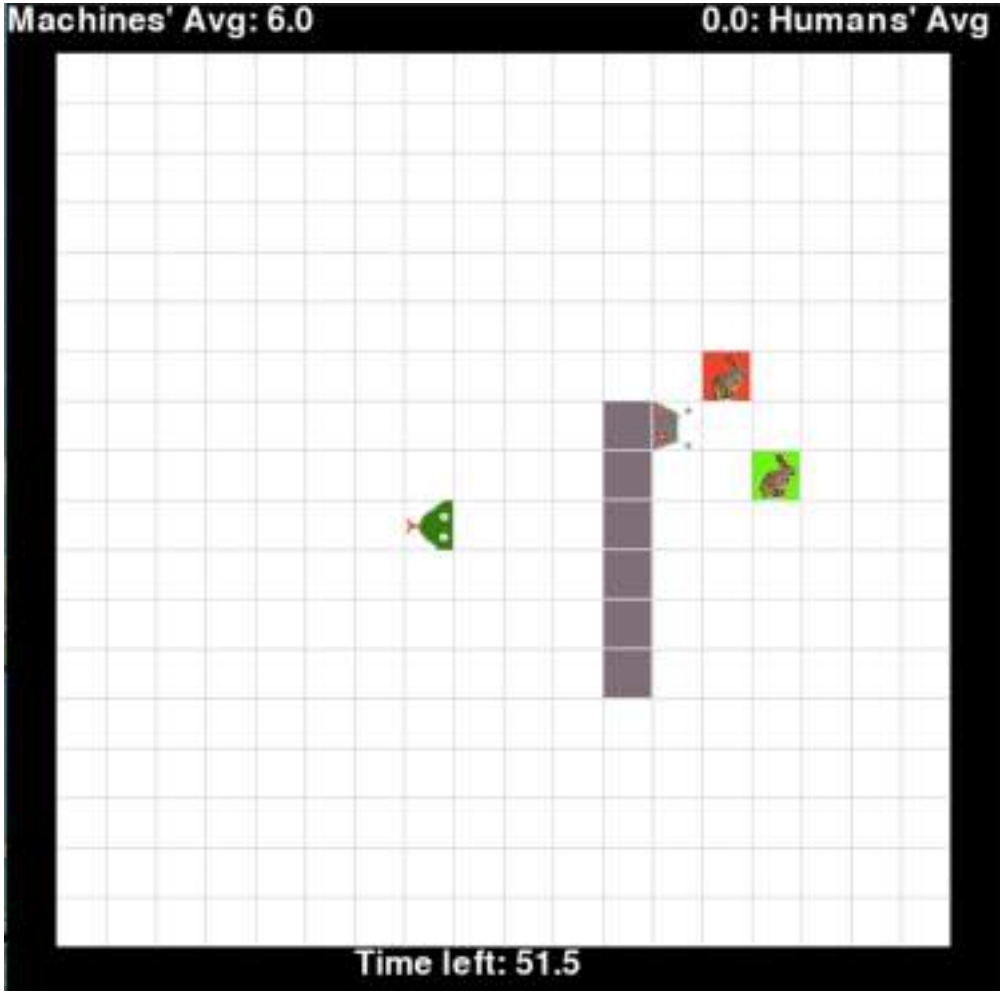
The Game played is…



DATA SCIENCE RETREAT

The rules are…

1. Eat the green bunnies
2. The relevant score is Food/ Death Average
3. 60 Seconds time limit



DATA SCIENCE RETREAT