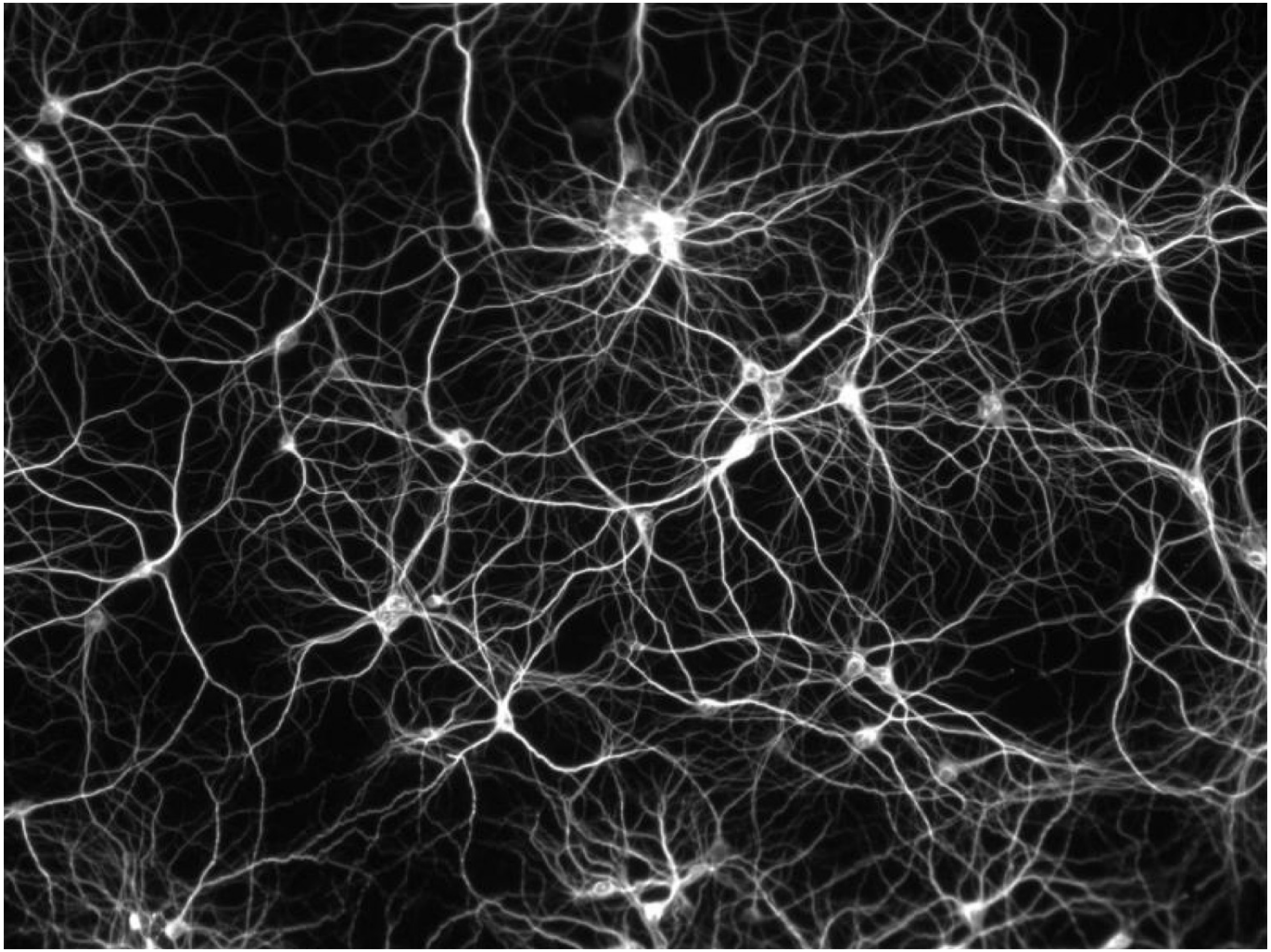
Neural Networks: Perceptron

Learning outcomes

- Describe the biological principles that inspired neural networks.
- Draw the diagram of the McCulloch and Pitts's neuron.
- Describe the di erence between zero, first, and second-order optimisation methods.
- Apply gradient descent to a given objective function.
- Choose an appropriate step size for gradient descent.

Our brain

UNIVERSITY OF LEEDS

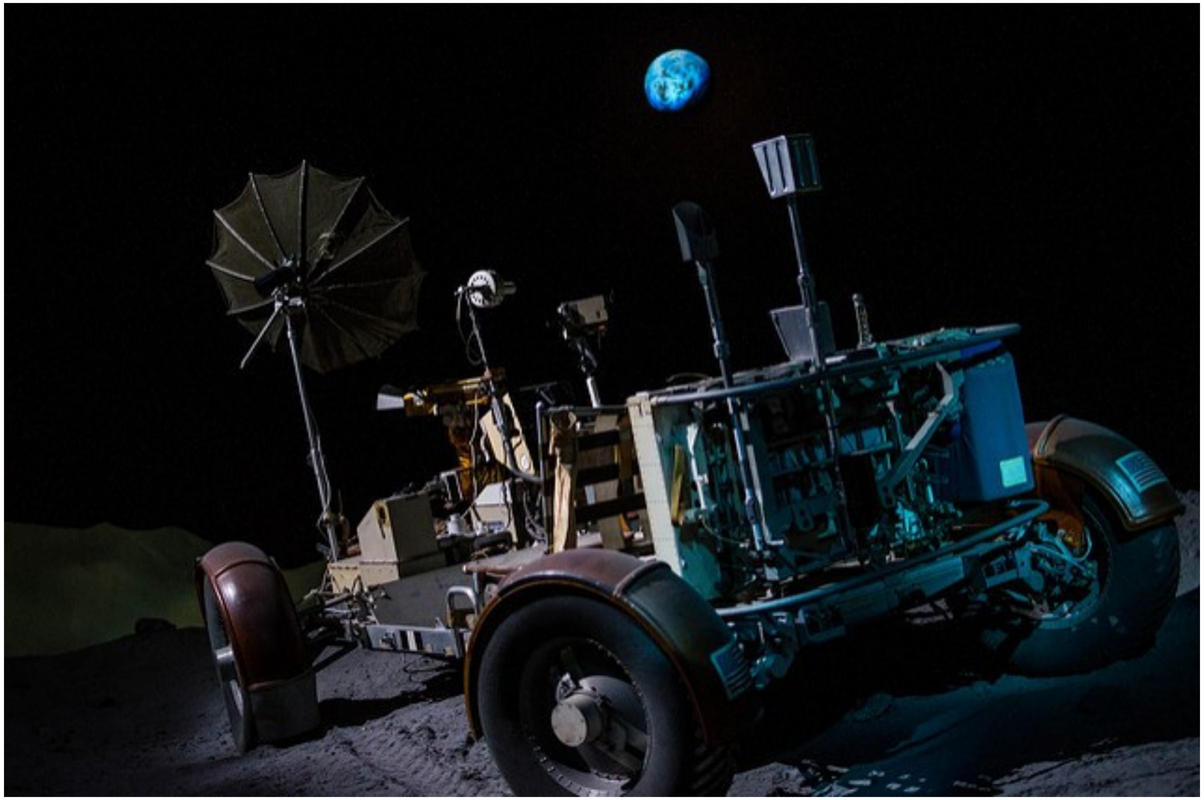


Neurons

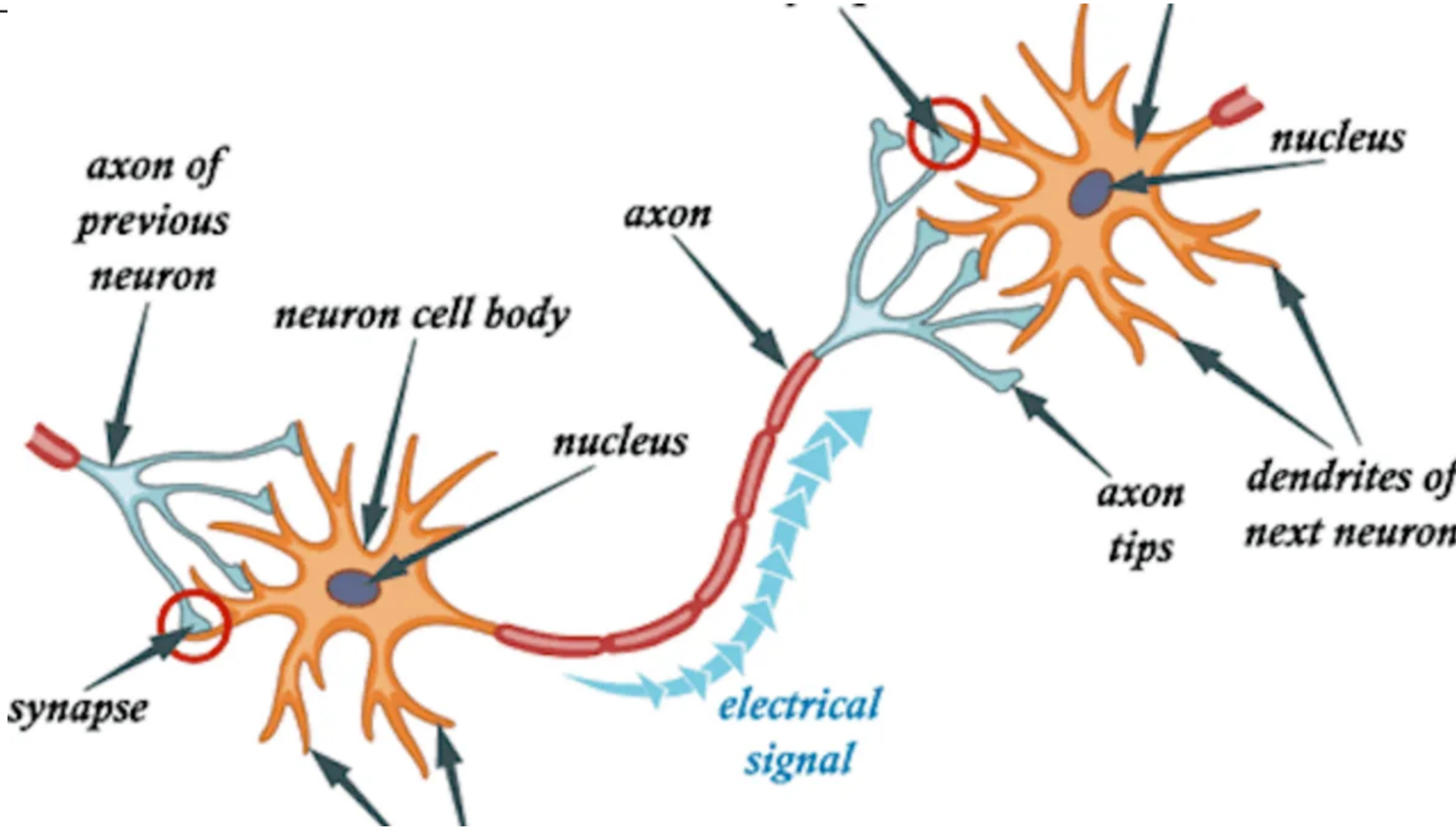
UNIVERSITY OF LEEDS

10^{11} neurons 10^{15} synapses

Enough fibres to cover the distance to the moon and back



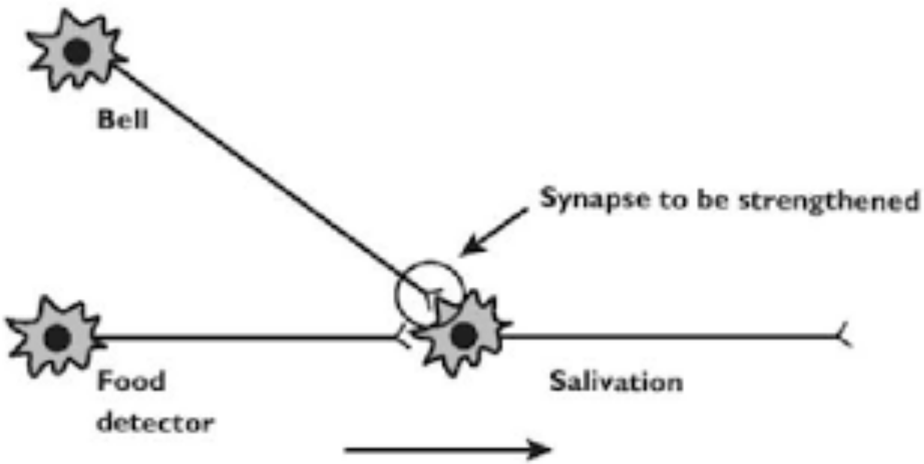
Firing



axon of previous neuron
neuron cell body
nucleus
axon
nucleus
synapse
electrical signal
axon tips
dendrites of next neuron

Hebbian learning

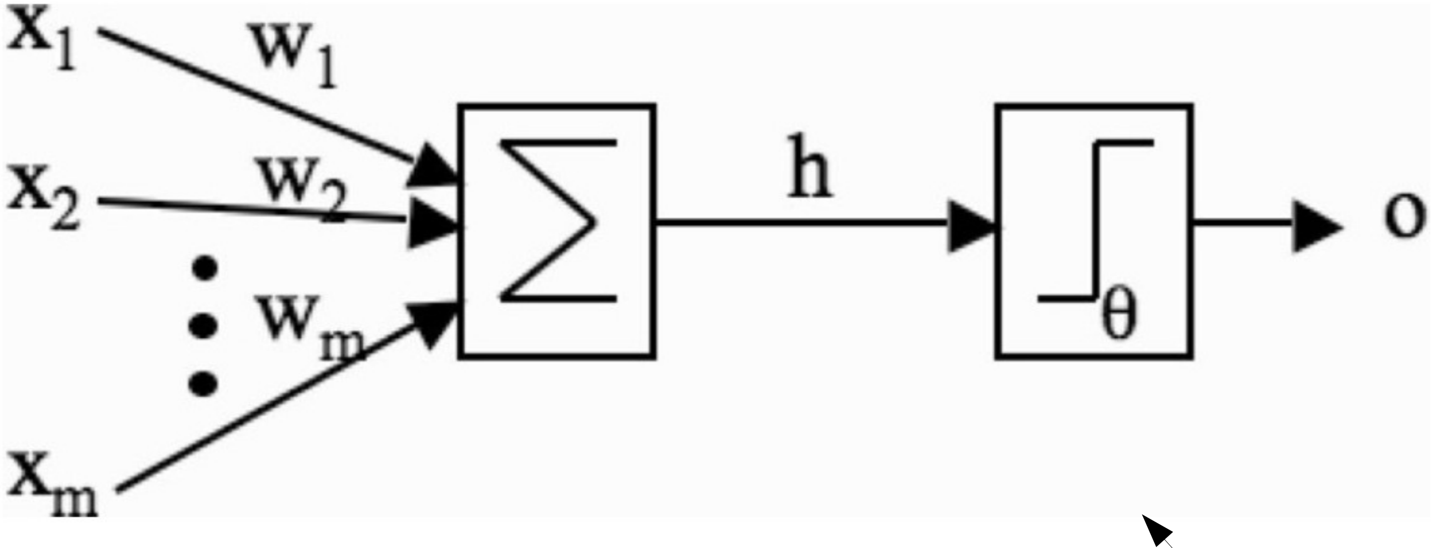
UNIVERSITY OF LEEDS



When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased

Donald Hebb (1949)

McCulloch and Pitts Neuron



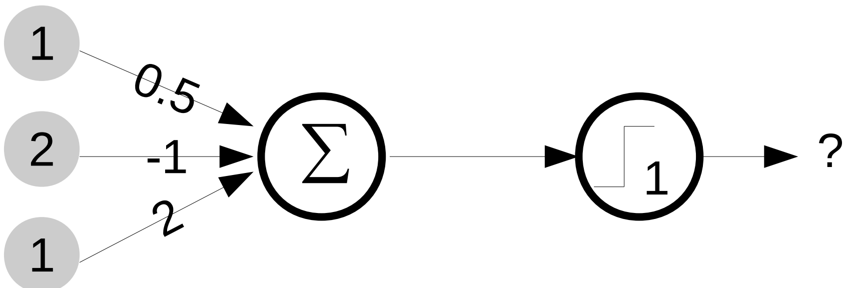
Activation function

$$h_w(\boldsymbol{x}) = \sum_i w_i x_i = \boldsymbol{w} \cdot \boldsymbol{x}$$

$$o(h_w) = \begin{cases} 1 & \text{if} \quad h_w > \theta \\ 0 & \text{if} \quad h_w \leq \theta \end{cases}$$

Question

What is the output of the following neuron?



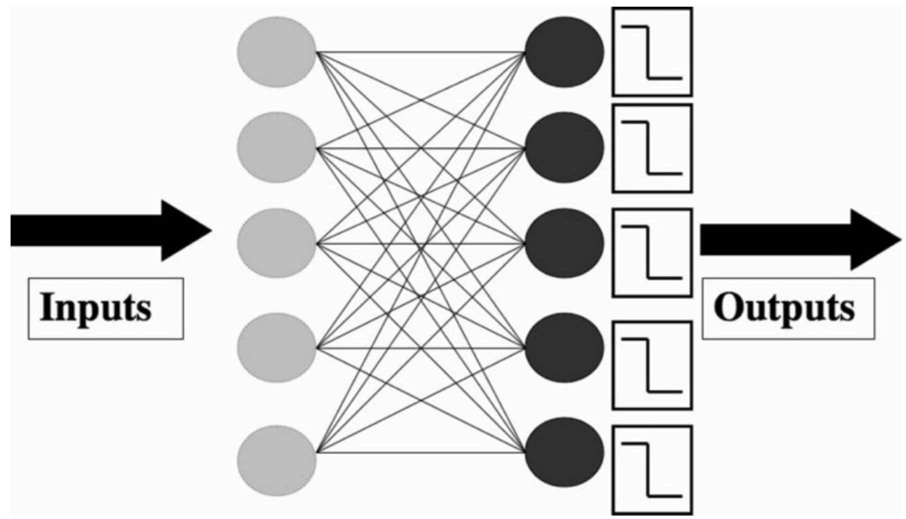
1: -1
2: 0
3: 1

https://www.strawpoll.me/16615211

Perceptron

UNIVERSITY OF LEEDS

The Perceptron network, consisting of a set of input nodes (left) connected to McCulloch and Pitts neurons using weighted connections.



We want the Perceptron to learn to reproduce a particular target, that is, a pattern of firing and non-firing neurons for the given input.

Training the perceptron

Learning happens through **optimisation**.

We define an error function, and then an optimisation algorithm finds the parameters that obtain the minimum error.

For example, the error function is the total number of mistakes:

$$E(\boldsymbol{X}) = \sum_{\vec{x}_n \in \boldsymbol{X}} |y_n - t_n|$$

Where y_n is the output of the perceptron on point n, and $t_n \in \{0,1\}$ is the desired classm, and $\boldsymbol{X}$ is the dataset.

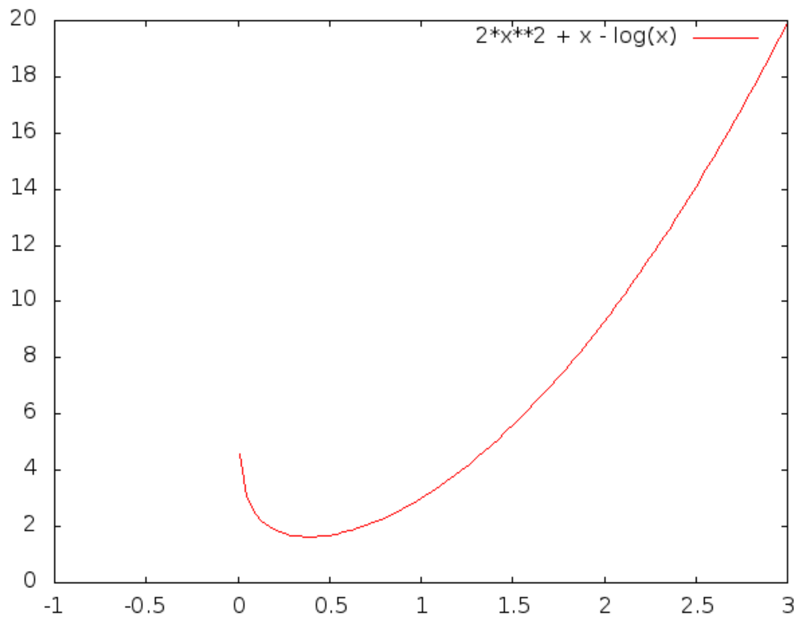
UNIVERSITY OF LEEDS

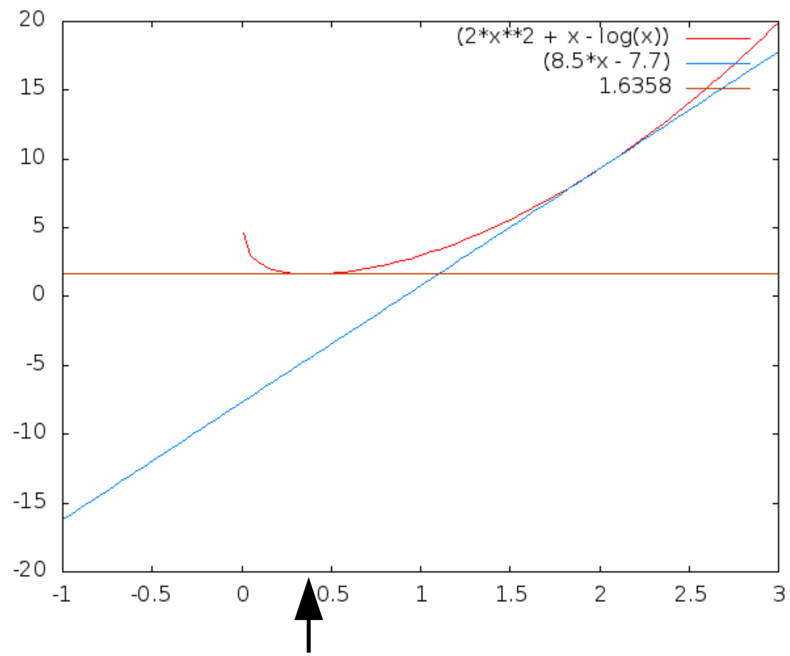
Optimisation

Goal

UNIVERSITY OF LEEDS

Find the minimum point of a given function:





The minimum is at 0.39

Gradient descent

UNIVERSITY OF LEEDS

First order: gradient descent

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

step parameter

Second order: Newton's method

$$f(x_n + \Delta x) \approx f(x_n) + f'(x_n)\Delta x + \frac{1}{2} f''(x_n)\Delta x^2$$

Taylor's expansion

$$\frac{\partial}{\partial \Delta x} f(x_n + \Delta x) = f'(x_n) + f''(x_n)\Delta x = 0$$

Optimal step

$$\Delta x = \frac{-f'(x_n)}{f''(x_n)}$$

Many dimensions: $x_{t+1} = x_t - H^{-1}\big|_{x_n} \nabla f$

Question

The current point is $<1,0>$, compute the next point following gradient descent on the function $f(x,y) = x^3 + 2y^2 - y$ with step size 0.1.

Question

UNIVERSITY OF LEEDS

We want to compute: $x_{t+1} = \langle 1,0 \rangle - 0.1 \nabla f(x_t)$

$\nabla f = \langle 3x^2, 4y - 1 \rangle$ Evaluated in <1,0> is <3,-1>

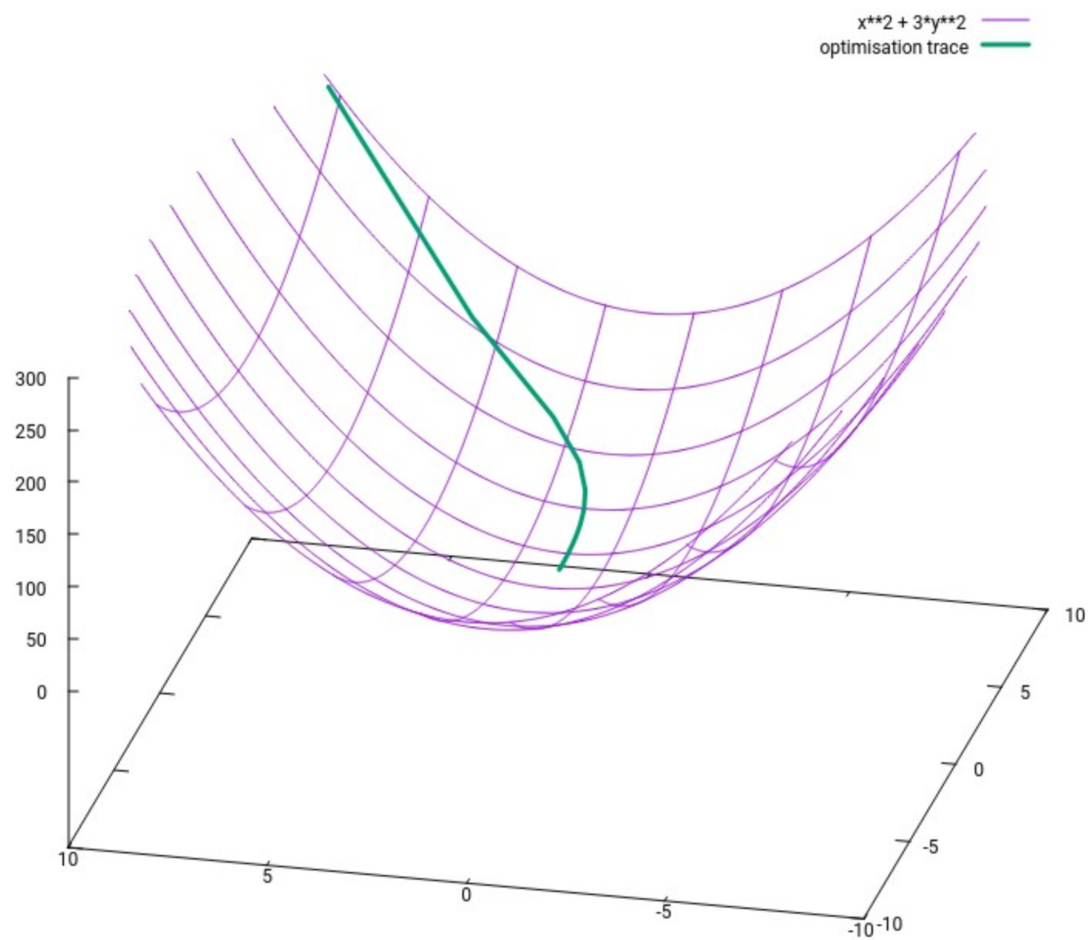
$$x_{t+1} = \langle 1,0 \rangle - 0.1 \cdot \langle 3, -1 \rangle = \langle 0.7, 0.1 \rangle$$

$f(1,0) = 1$

Our solution has improved!

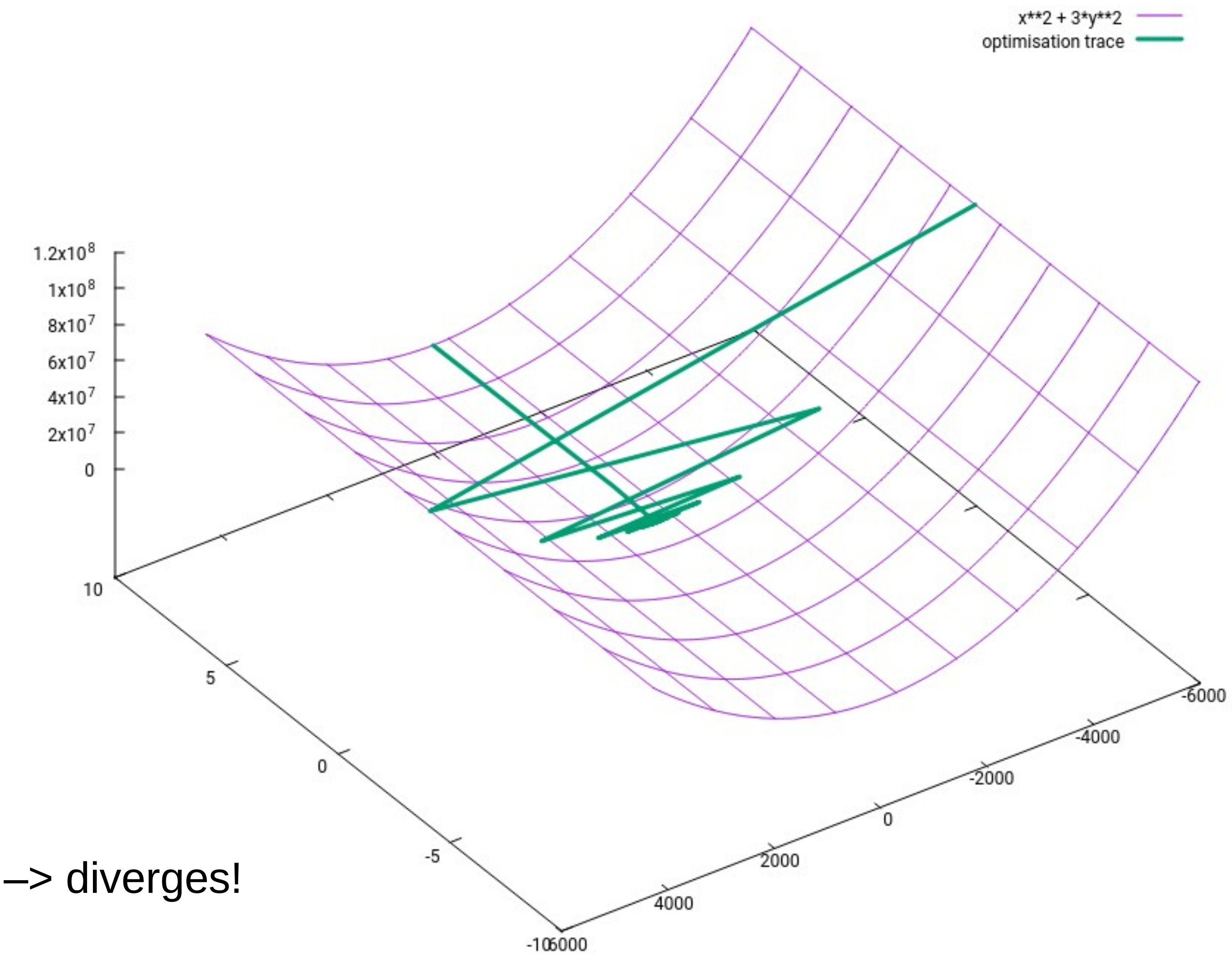
$f(0.7, 0.1) = 0.263$

In 3D



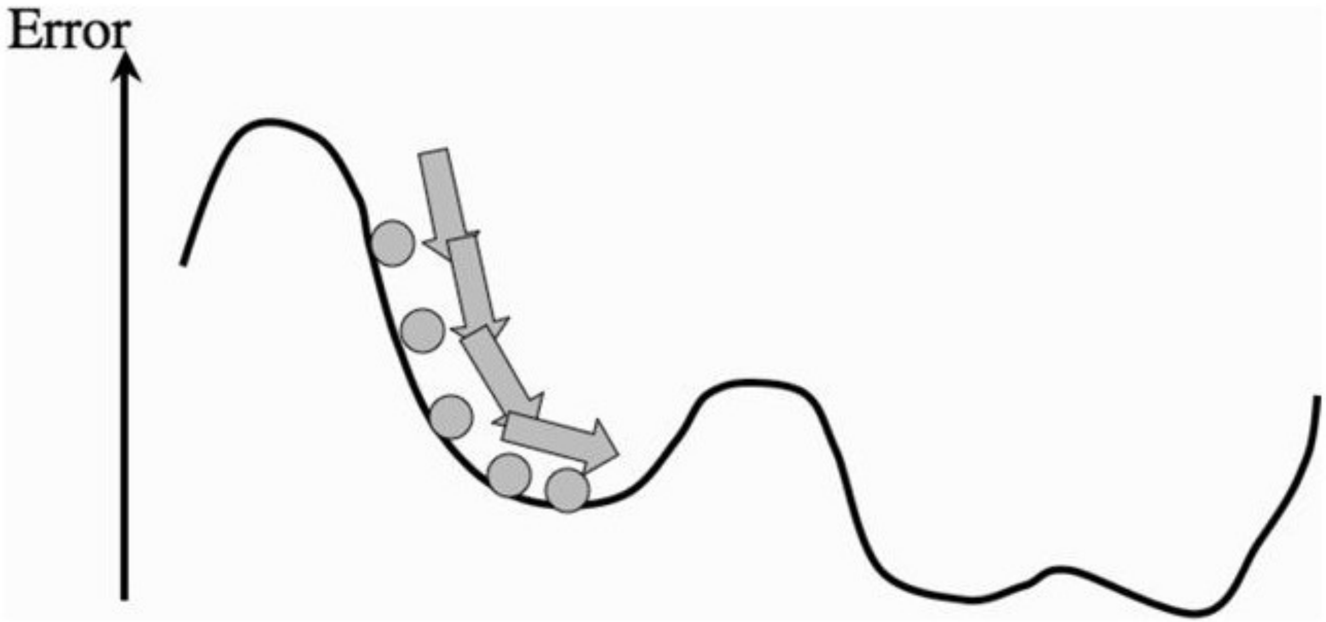
Step size: 0.1

In 3D



Step size: 0.14 –> diverges!

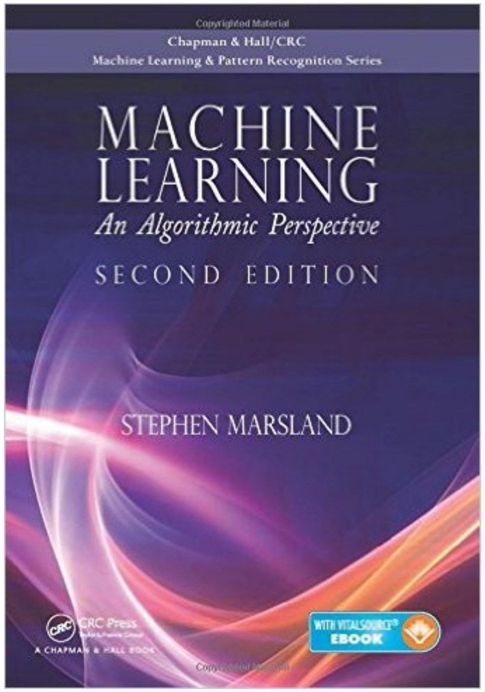
Local Minima



Conclusion

Learning outcomes

- Describe the biological principles that inspired neural networks.
- Draw the diagram of the McCulloch and Pitts’s neuron.
- Describe the di erence between zero, first, and second-order optimisation methods.
- Apply gradient descent to a given objective function.
- Choose an appropriate step size for gradient descent.



Chapter 3, up to 3.3
Section 9.0, 9.1