Class: Machine Learning

Elements of Local Optimisation

Instructor: Matteo Leonetti

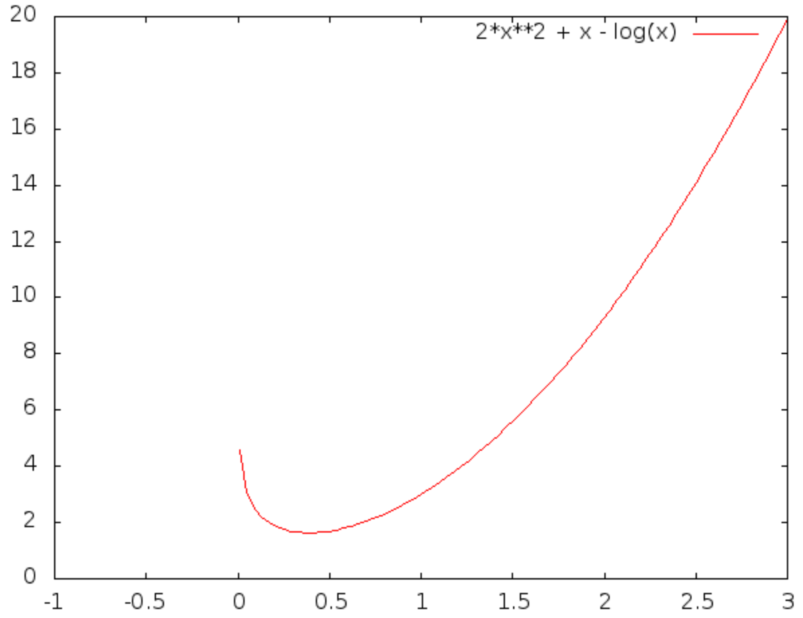
Learning outcomes

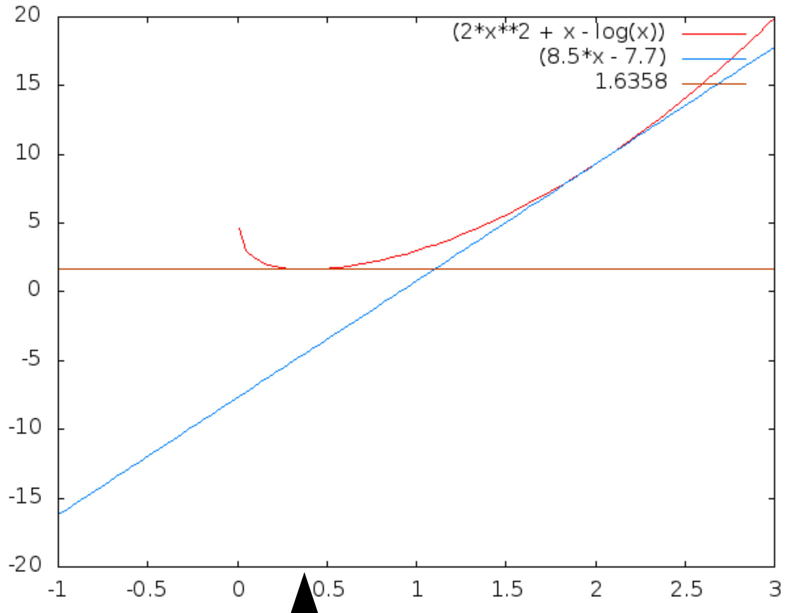
- Describe the difference between zero, first, and second-order optimisation methods.
- Apply gradient descent to a given objective function.
- Choose an appropriate step size for gradient descent.

Goal

UNIVERSITY OF LEEDS

Find the minimum point of a given function:





The minimum is at 0.39

Local Methods

UNIVERSITY OF LEEDS

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

UNIVERSITY OF LEEDS

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

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>

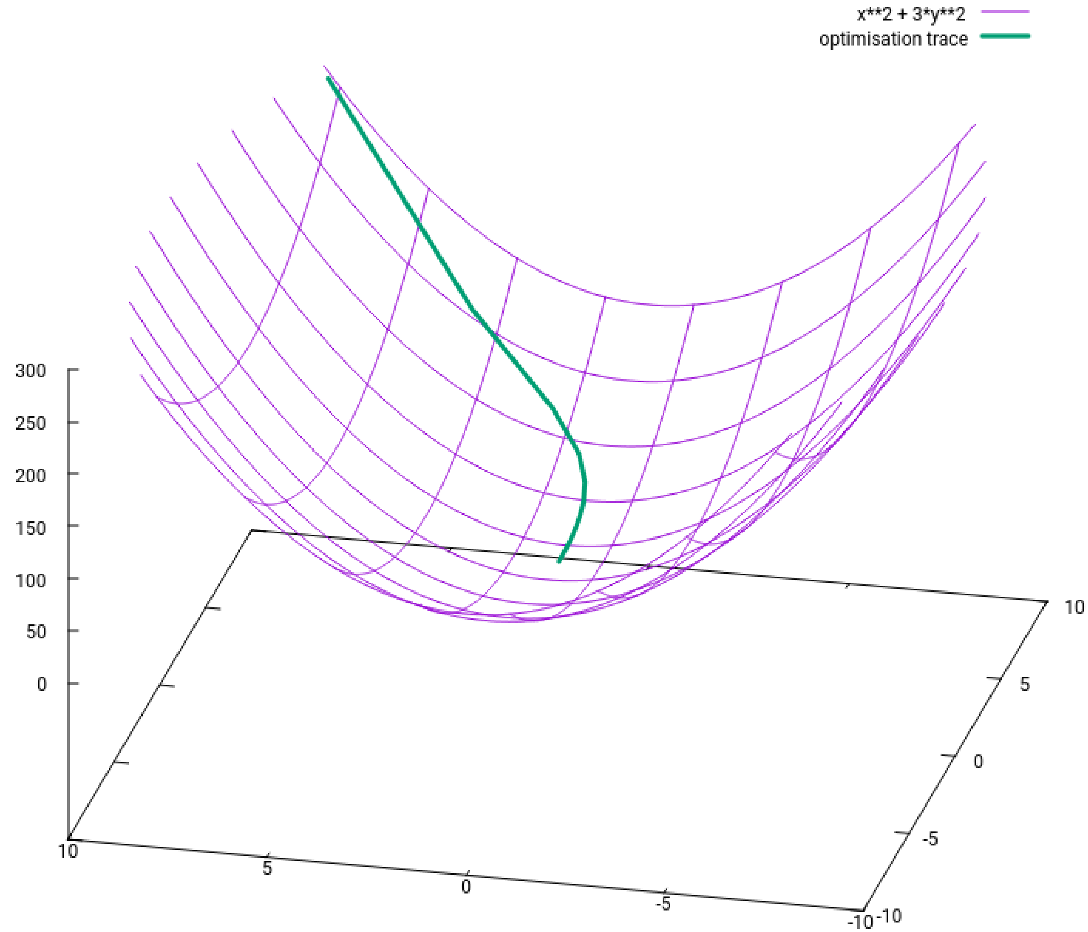
$$x_{t+1} = \langle 1,0 \rangle - 0.1 \cdot \langle 3, -1 \rangle = \langle 0.7, 0.1 \rangle$$

$f(1,0) = 1$

Our solution has improved!

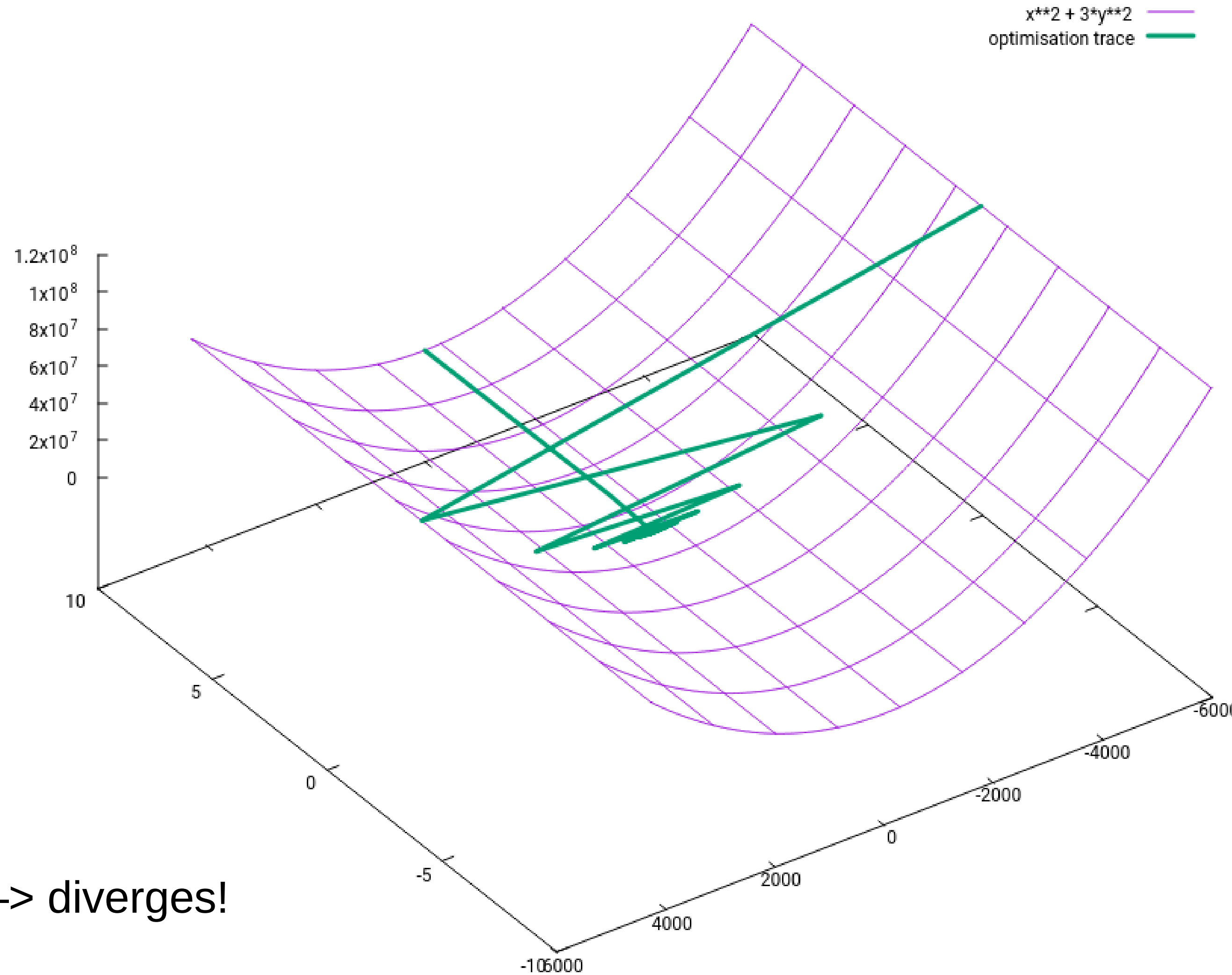
$f(0.7, 0.1) = 0.263$

In 3D



Step size: 0.1

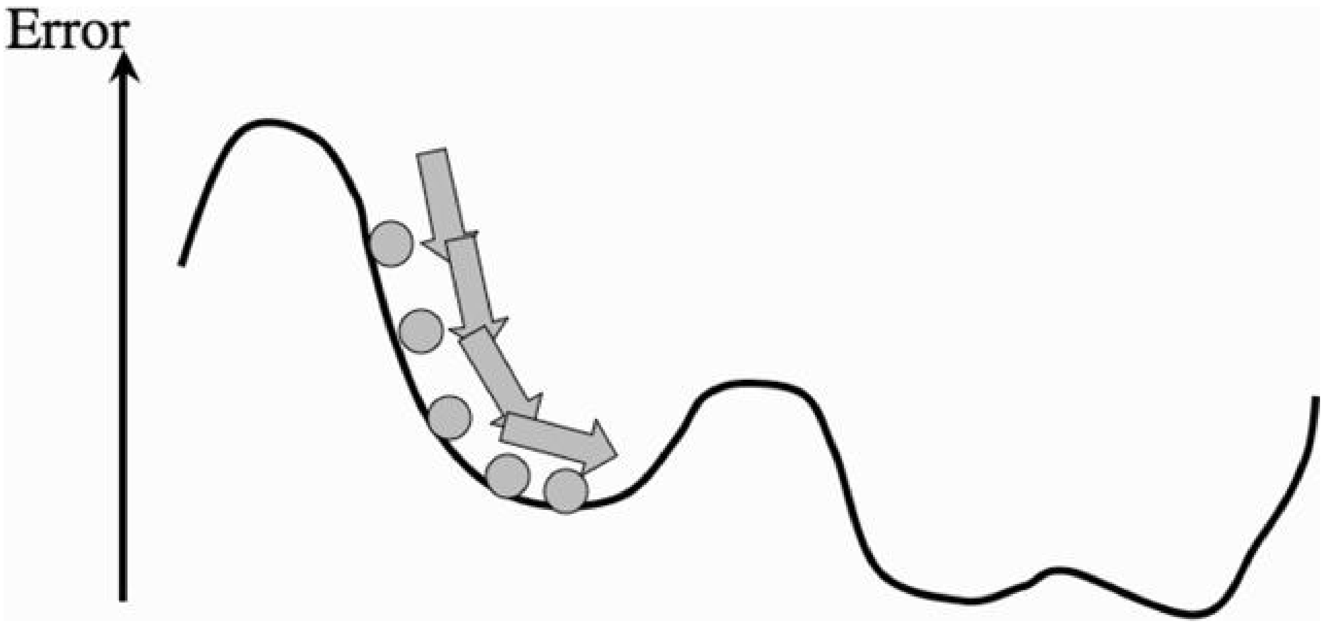
In 3D



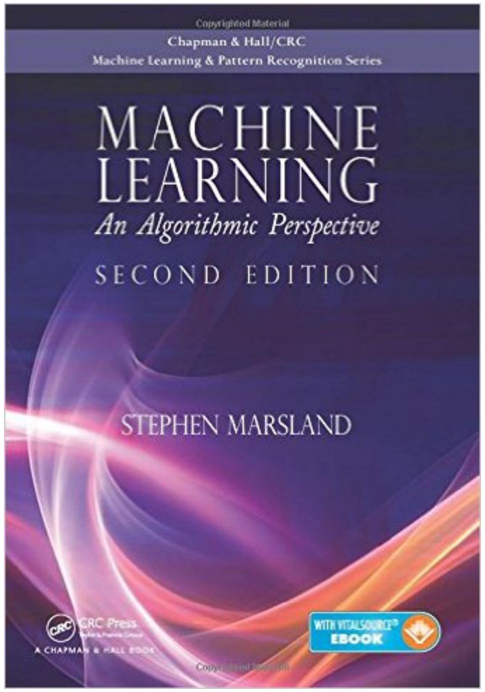
Step size: 0.14 –> diverges!

Local Minima

UNIVERSITY OF LEEDS



Conclusion



Sections 9.0, 9.1