Introduction to Neural Networks

Johns Hopkins University
Engineering for Professionals Program
605-447.71/625-438.71
Dr. Mark Fleischer

Module 2.2: Mathematical Review-Differential Calculus

This Sub-Module Covers …

Review of differential calculus:

- derivatives, partial derivatives, gradients
- directional derivatives

The next sub-module covers:

- Calculus-based optimization methods and related material:
  - First order necessary conditions.
  - Second order sufficiency conditions.
  - Definition of convexity.

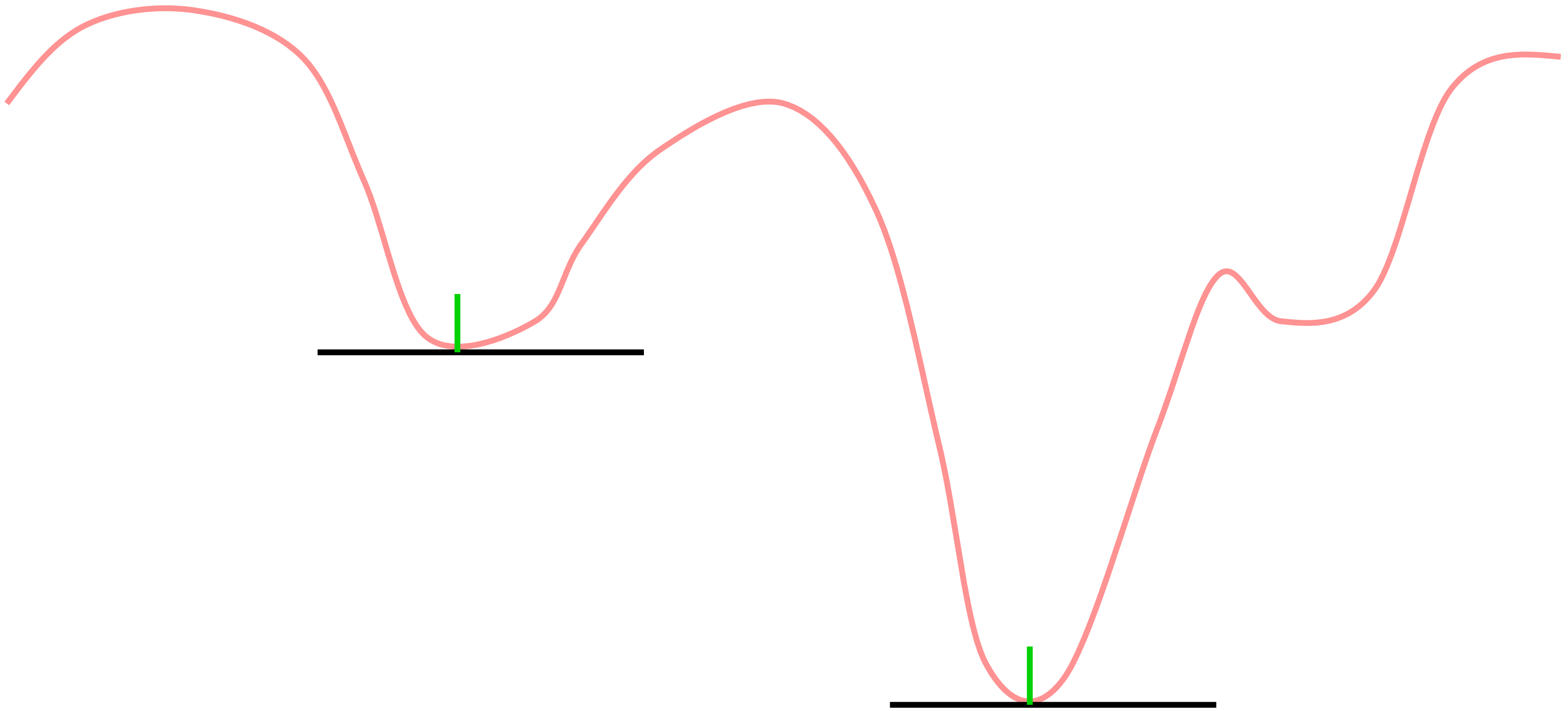
Optimization and Learning

Many types of optimization problems.

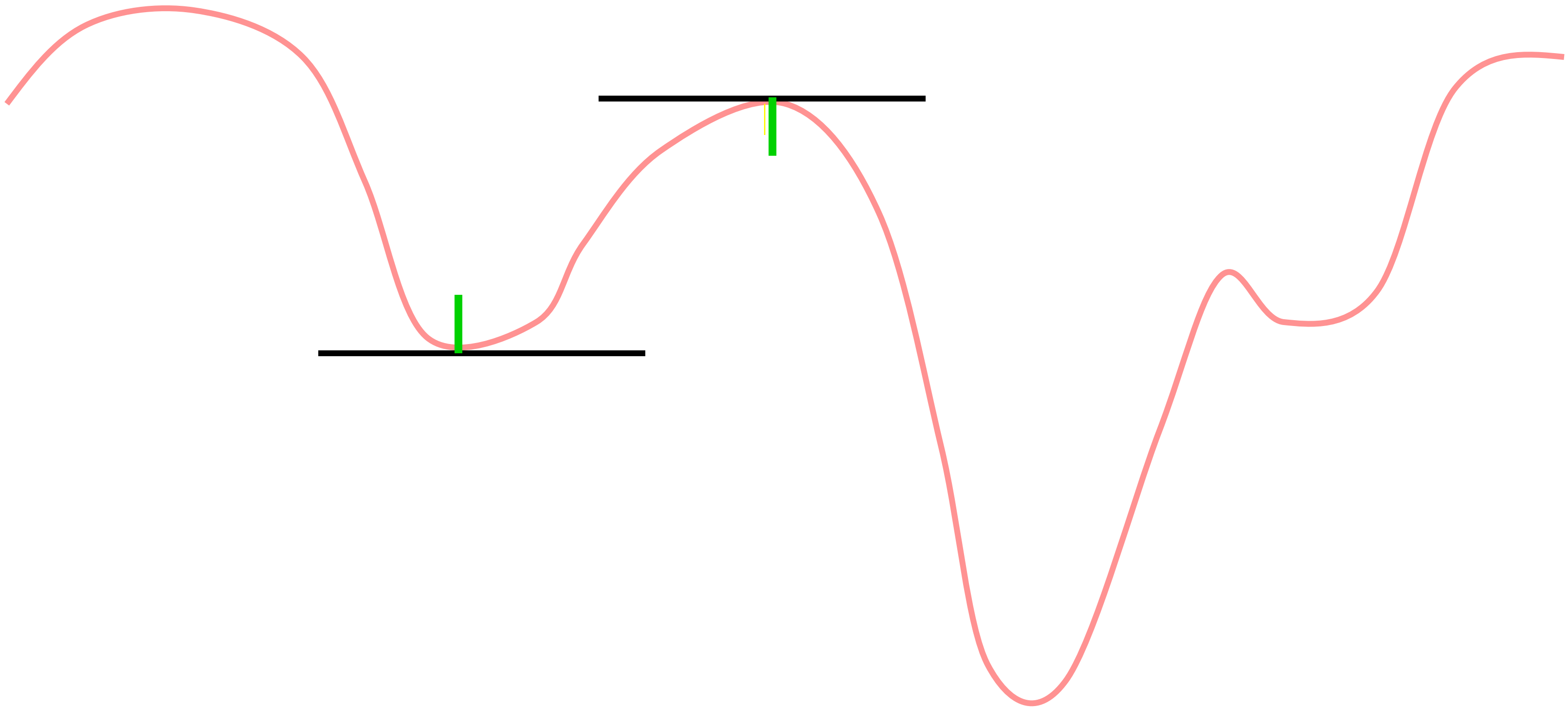
- For continuous functions, we use calculus-based methods.
- Problems with linear objective functions, linear constraint equations can be solved using Linear Programming methods
- Problems with non-linear objective functions, non-linear constraint equations can be solved using Non-Linear Programming methods.
- For now, we will not worry about constraints.

Training a neural network involves solving an optimization problem!

Calculus Based Optimization



Calculus Based Optimization



Derivatives

$$\frac{dy}{dx} = \frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$\frac{\partial f(x)}{\partial x_i} = \frac{f(x_1, \cdots, x_i + h, \cdots, x_n) - f(x_1, \cdots, x_i, \cdots, x_n)}{h}$$

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(x_1)}{\partial x_1}, \frac{\partial f(x_2)}{\partial x_2}, \cdots, \frac{\partial f(x_n)}{\partial x_n} \right)$$

Derivatives

$$f(x) = ax^n \Rightarrow \frac{df(x)}{dx} = f'(x) = nax^{n-1}$$

$$f(x_1, x_2) = ax_1 x_2^n \Rightarrow \frac{\partial f(x_1, x_2)}{\partial x_2} = nax_1 x_2^{n-1}$$

Directional Derivatives

- In partial derivatives, we consider the slope of a surface in the direction along one variable axis.
- In directional derivatives, we consider the slope of a surface in a specified direction.

Important for training neural networks!

Directional Derivatives

$$\frac{d f(x, y, z)}{d\mathbf{d}} = \lim_{t \to 0^+} \frac{f(x + td_1, y + td_2, z + td_3) - f(x, y, z)}{t}$$

Each independent variable x, y and z is perturbed by a component of the direction vector **d** and multiplied by a factor *t.*

Directional Derivatives

$$\begin{aligned}
\frac{d\, f(x,y,z)}{d\mathbf{d}} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} \\
&= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) \\
&= \nabla f(x,y,z) \cdot \mathbf{d}
\end{aligned}$$

Directional Derivatives