Gradient Descent

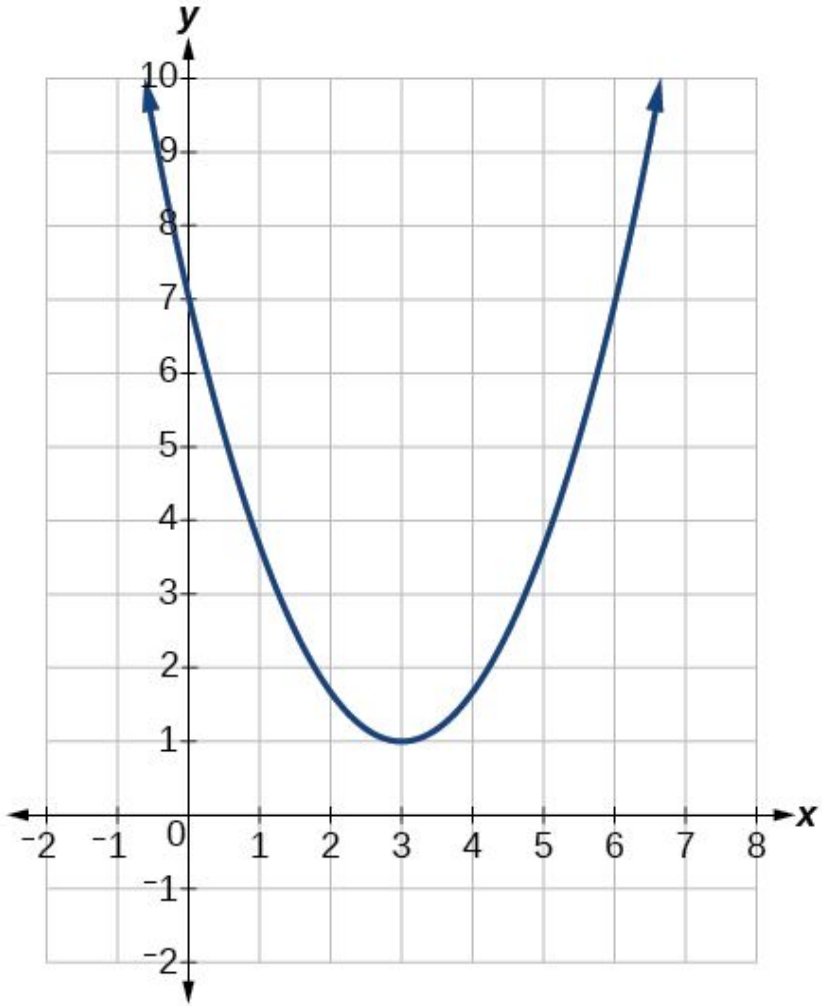
Algorithms Club 10.7.24

Wanderer above the Sea of Fog

By: Caspar David Friedrich

Basics

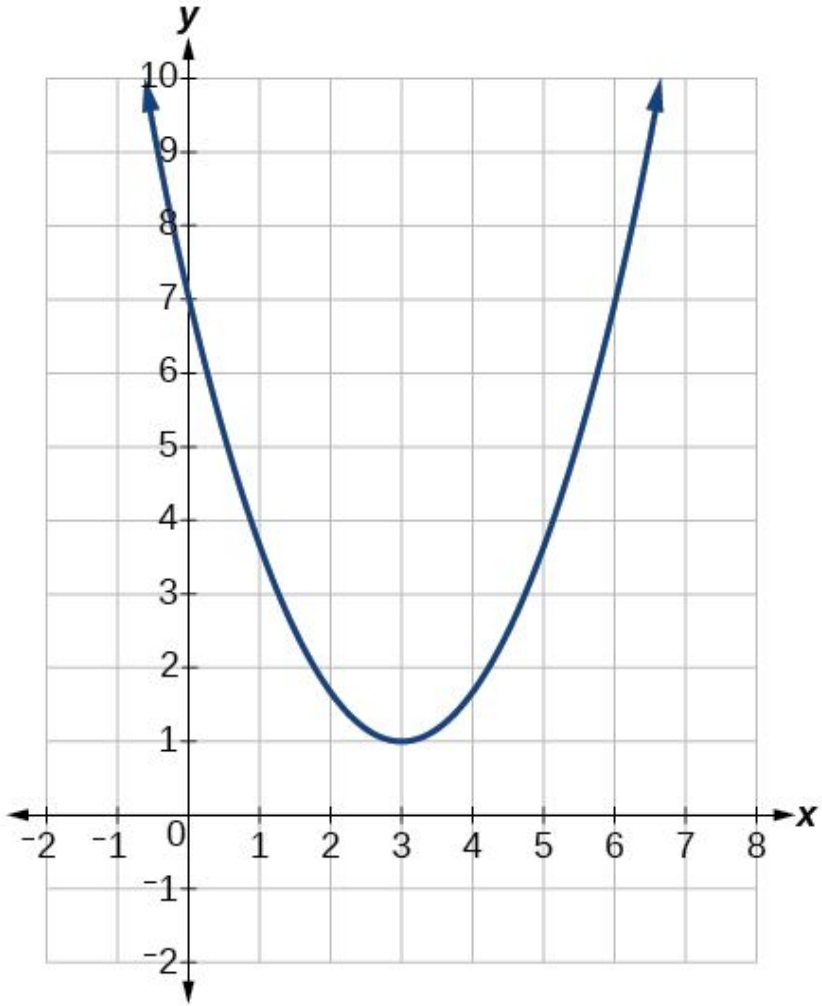
- Minimize a loss/cost function



Basics

- Minimize a loss/cost function

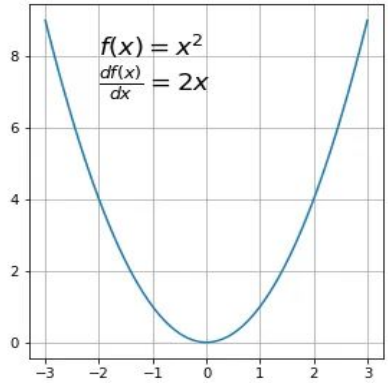
- Function Requirements

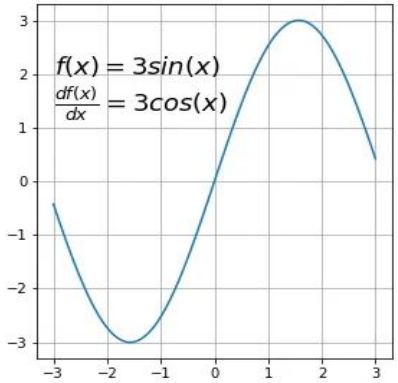


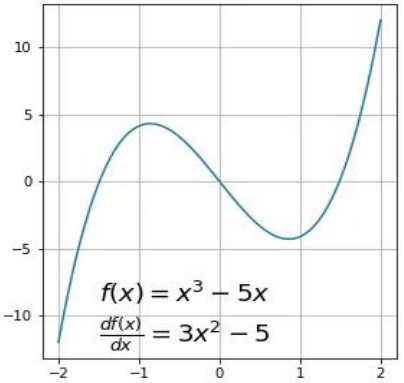
Basics

- Minimize a loss/cost function

- Function Requirements
  1. Differentiable



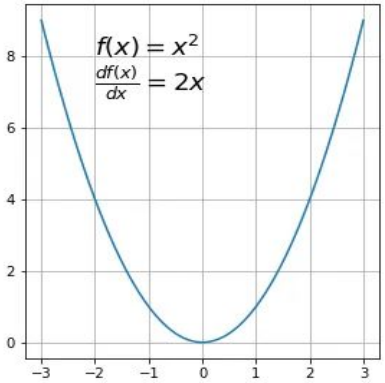


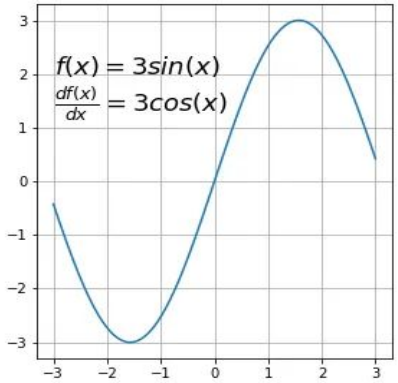


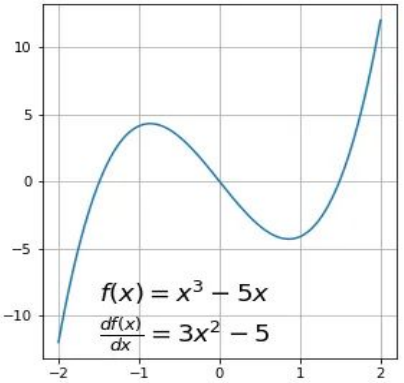
Basics

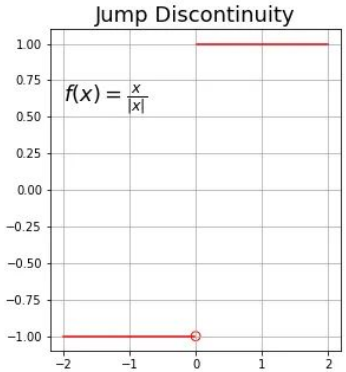
- Minimize a loss/cost function

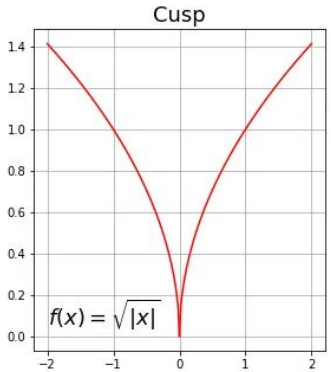
- Function Requirements
  1. Differentiable

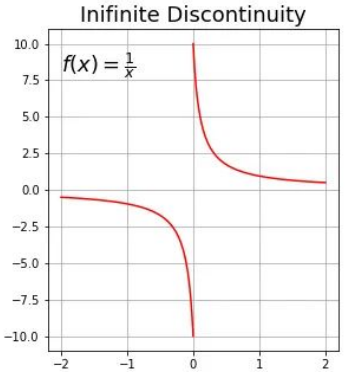








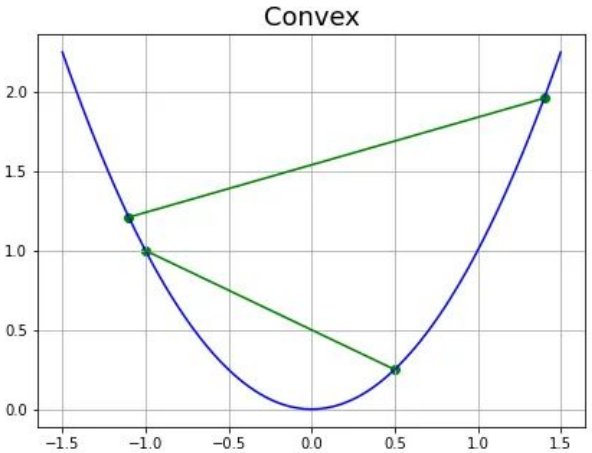


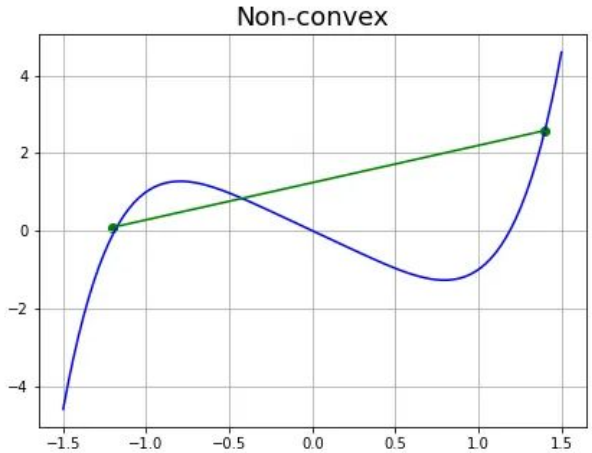


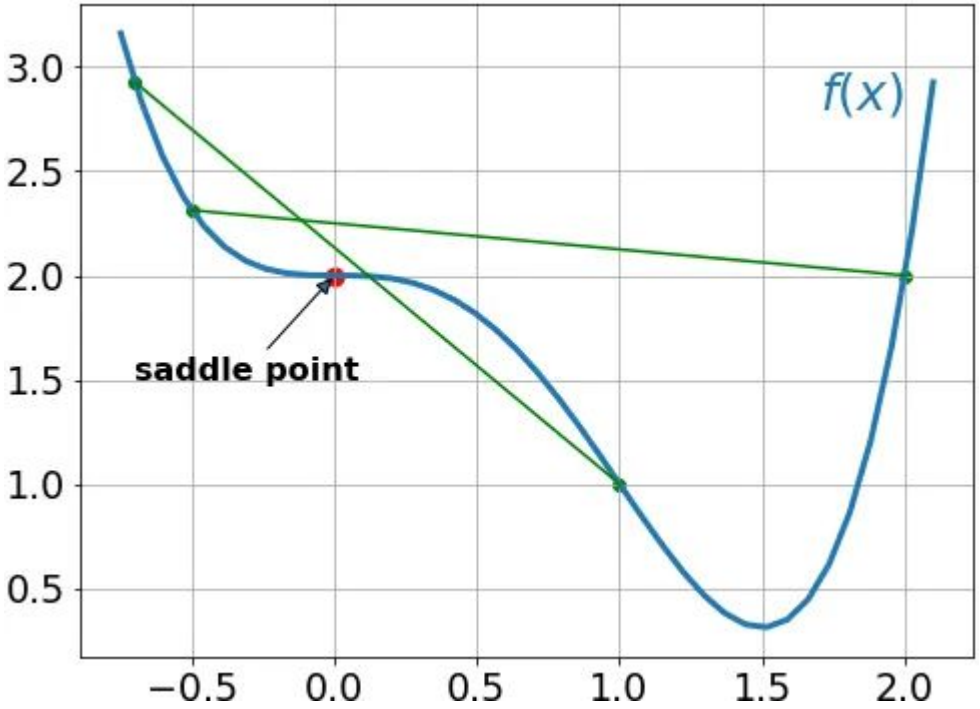
Basics

- Minimize a loss/cost function

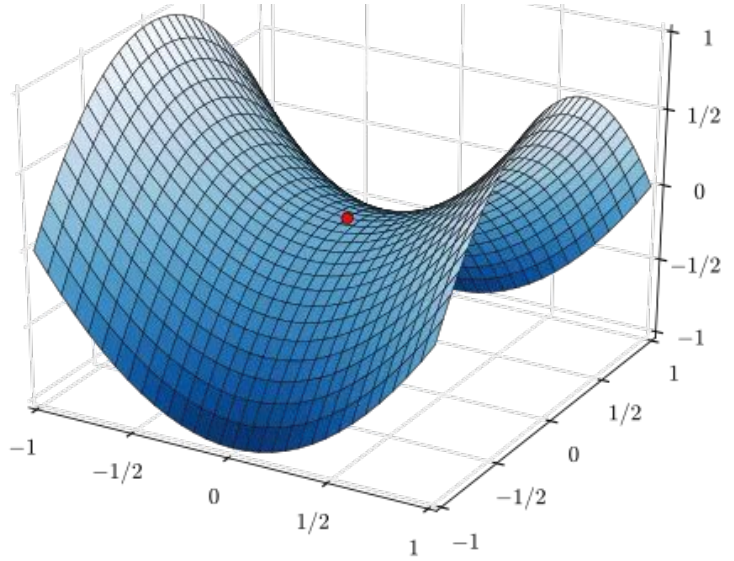
- Function Requirements
  1. Differentiable
  2. Convex





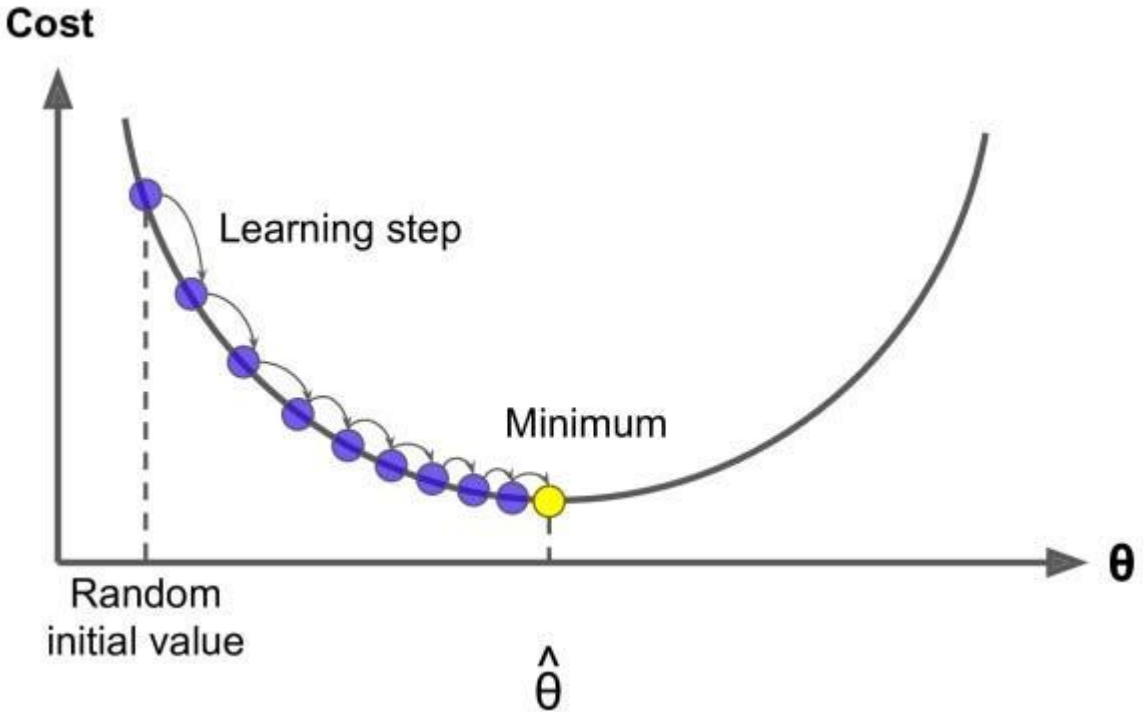


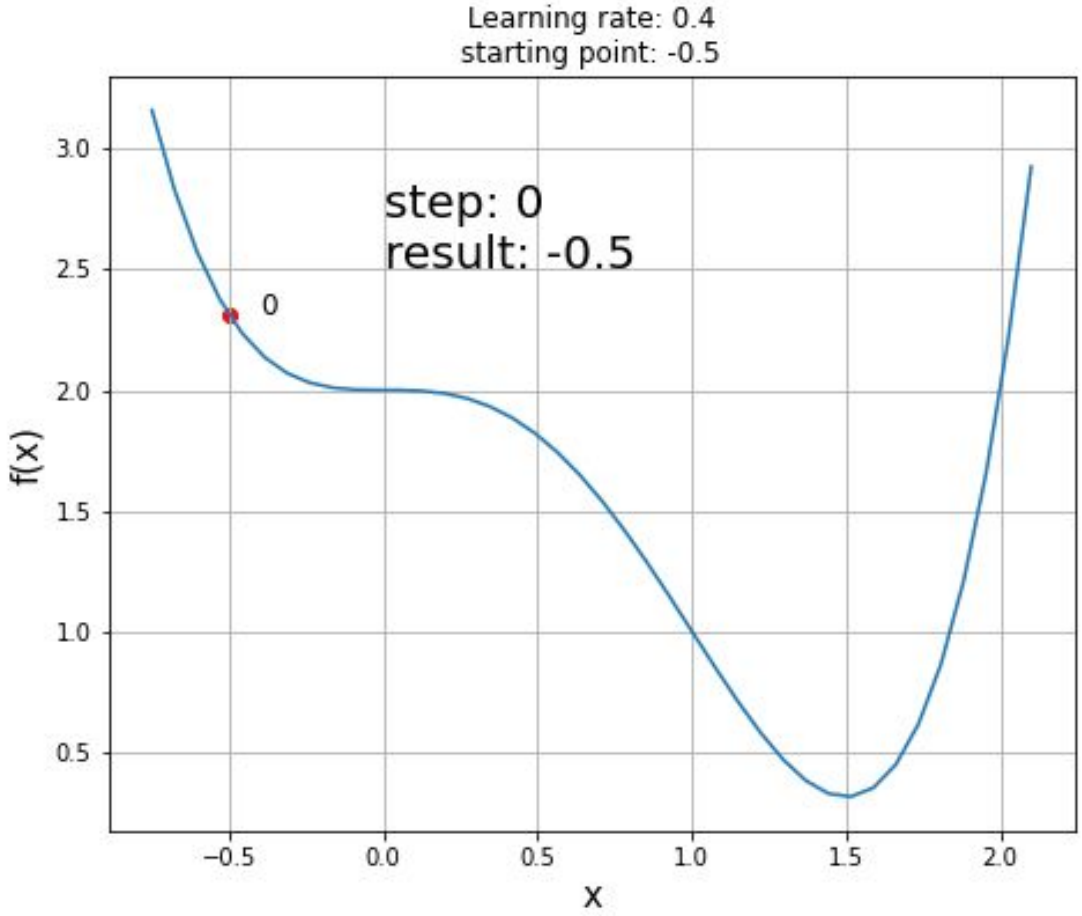
Local Minima and Saddle Points



More Basics

- Step-size

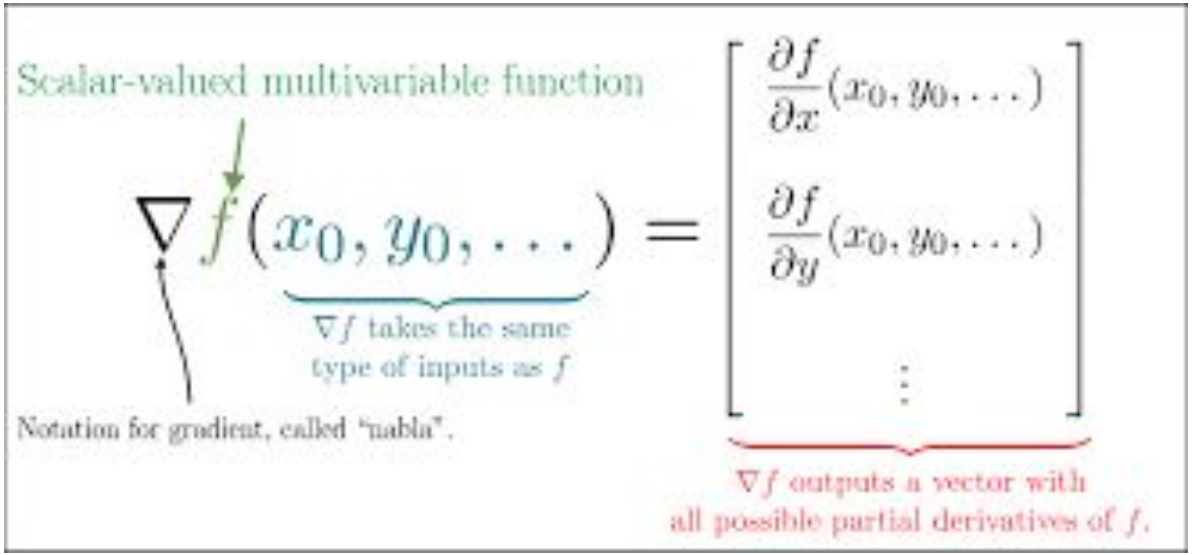



Learning rate: 0.4
starting point: -0.5
step: 0
result: -0.5
0
f(x)
x
3.0
2.5
2.0
1.5
1.0
0.5
−0.5
0.0
0.5
1.0
1.5
2.0

More Basics

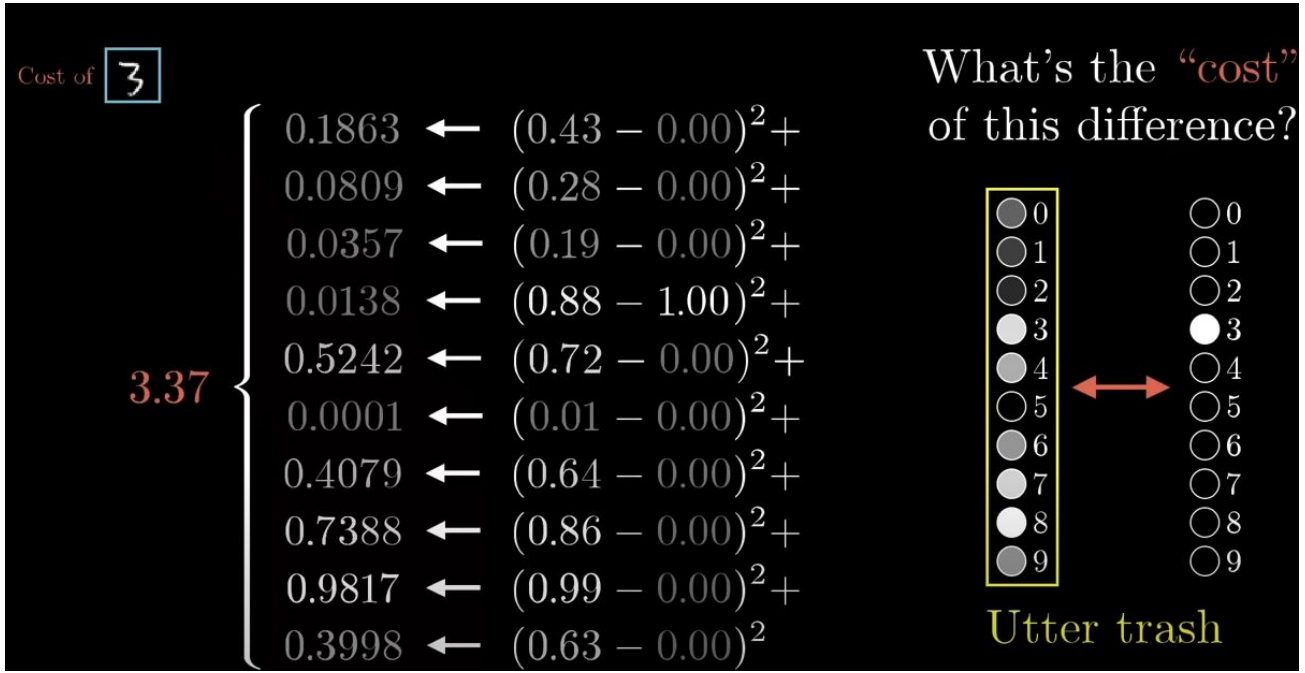
- Step-size

- Partial Derivatives



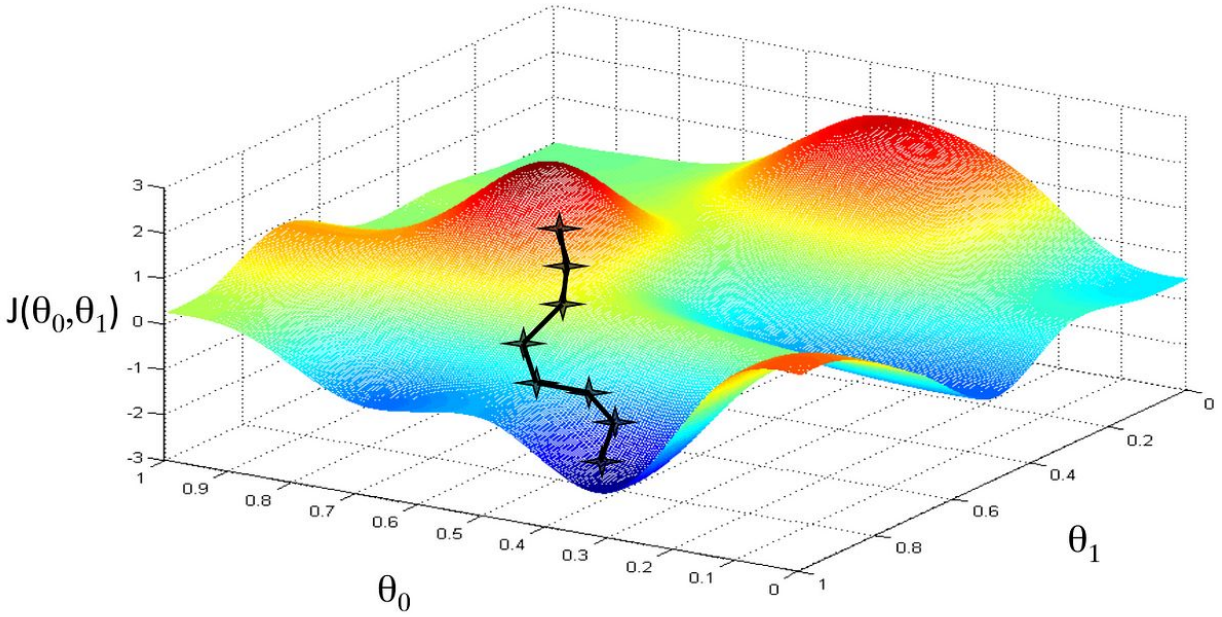
More Basics

- Step-size
- Gradient
- Partial Derivatives



More Basics

- Step-size
- Partial Derivatives
- Gradient



Pseudocode More

1. Choose a random starting point
2. Calculate the gradient at the point
3. Make a scaled step in the opposite direction of the gradient
4. Repeat until one of below criteria are met
    a. Max iterations reached
    b. Step size smaller than tolerance

References:
https://towardsdatascience.com/gradient-descent-algorithm-a-deep-dive-cf04e8115f21

- Khan partial derivatives
https://www.youtube.com/watch?v=dfvnCHqzK54

- 3Blue1Brown

https://www.youtube.com/watch?v=IHZwWFHWa-w

- Statquest

https://www.youtube.com/watch?v=sDv4f4s2SB8

- Blog post

https://colah.github.io/

- Distill
https://distill.pub/