

Gradient Descent

Algorithms Club 10.7.24

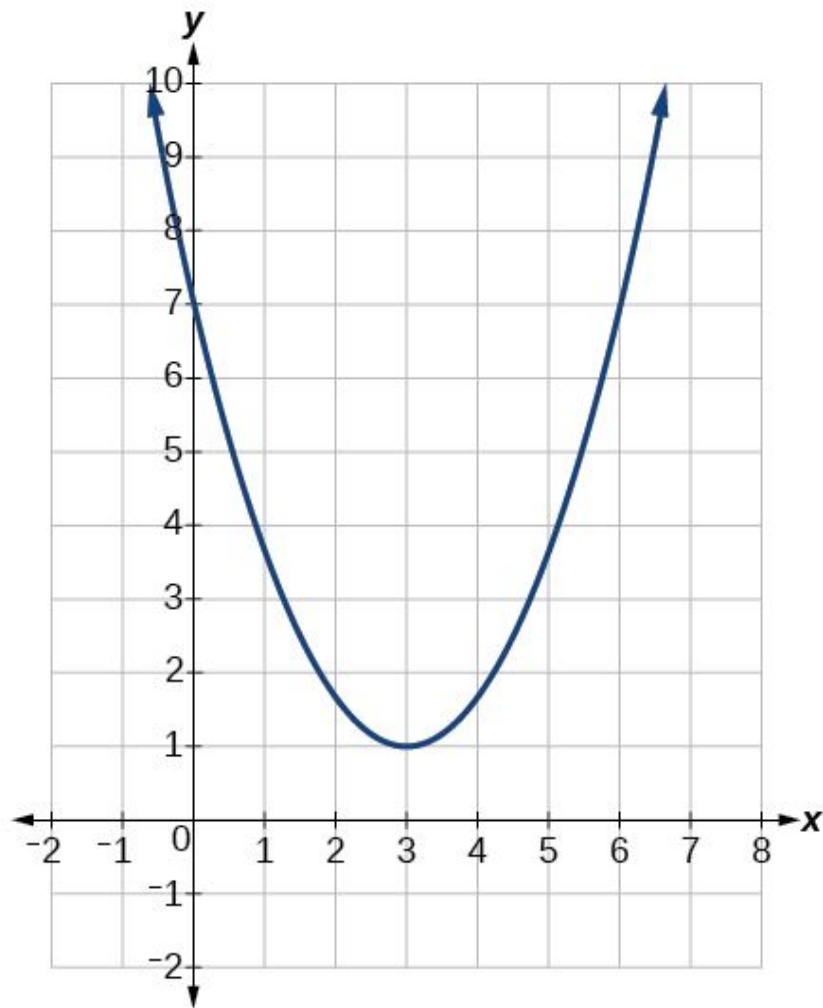
Wanderer above the Sea of Fog

By: Caspar David Friedrich



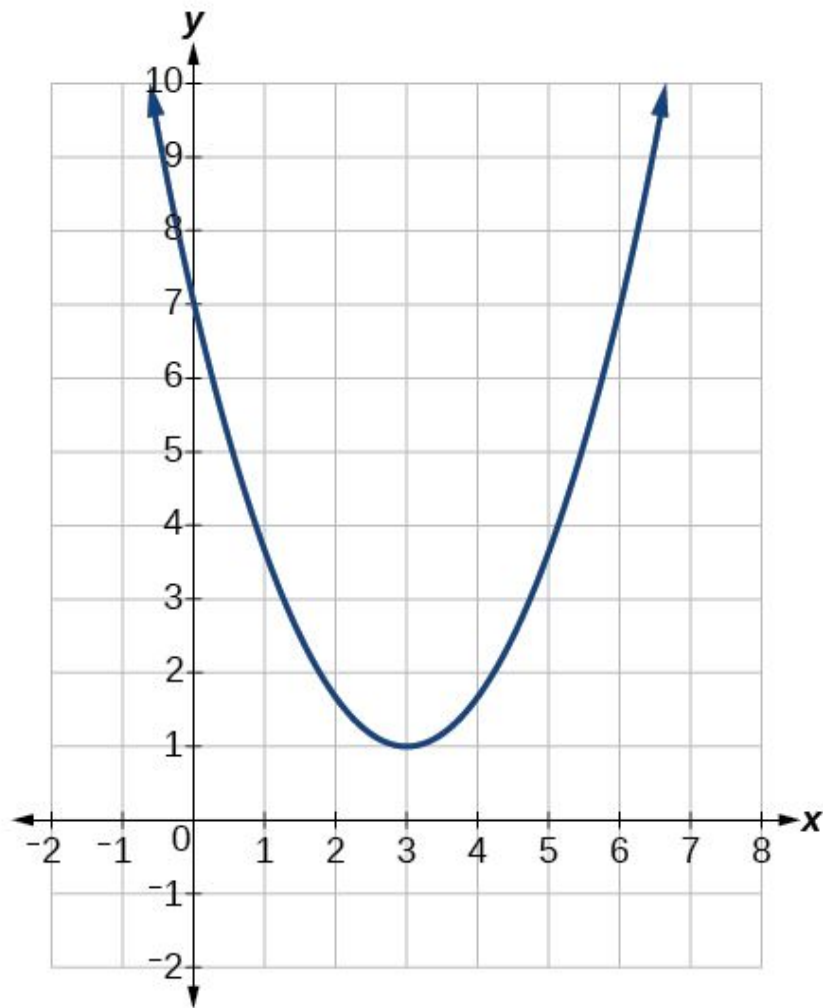
Basics

- Minimize a loss/cost function



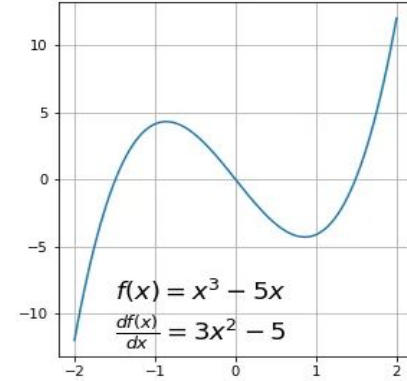
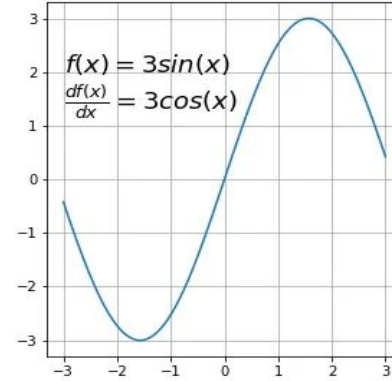
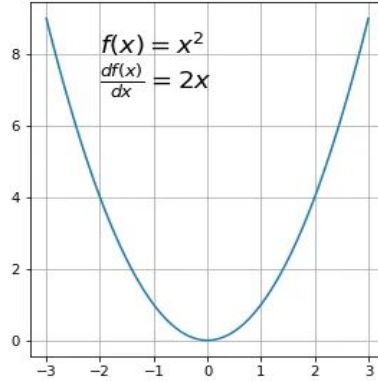
Basics

- Minimize a loss/cost function
- Function Requirements



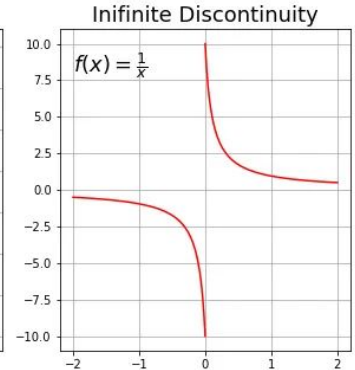
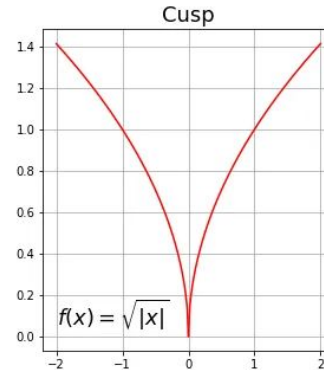
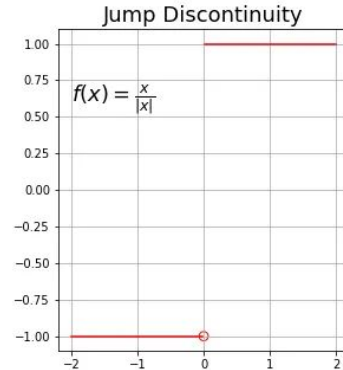
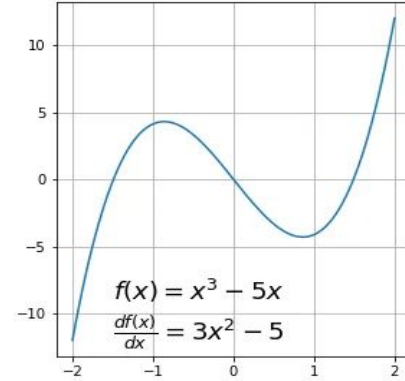
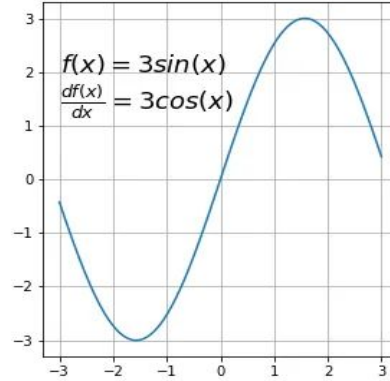
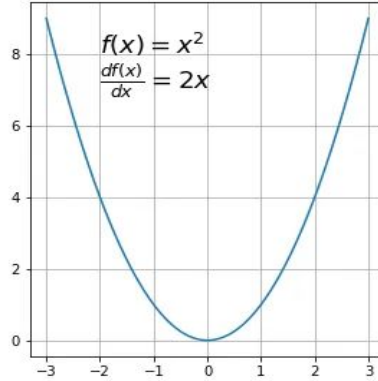
Basics

- Minimize a loss/cost function
- Function Requirements
 1. Differentiable



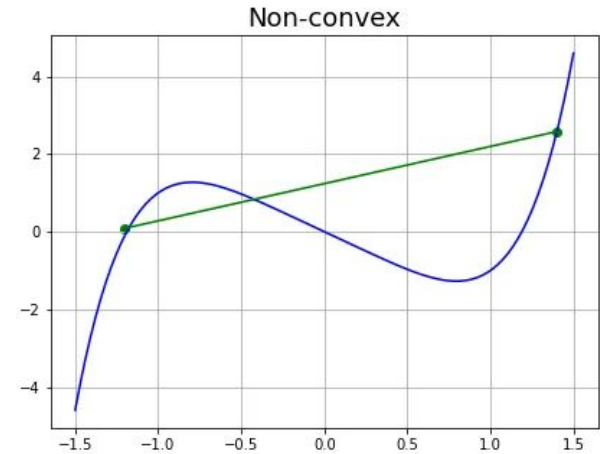
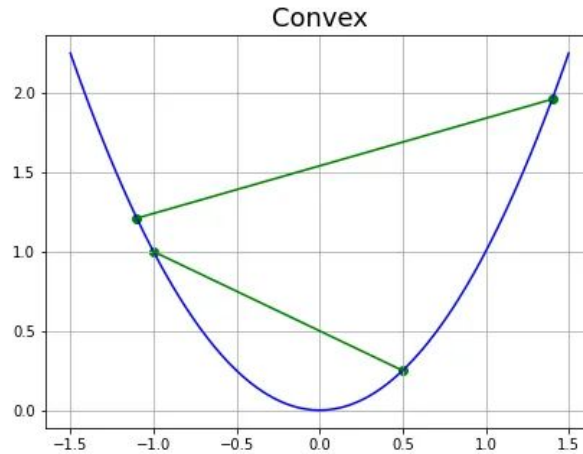
Basics

- Minimize a loss/cost function
- Function Requirements
 1. Differentiable

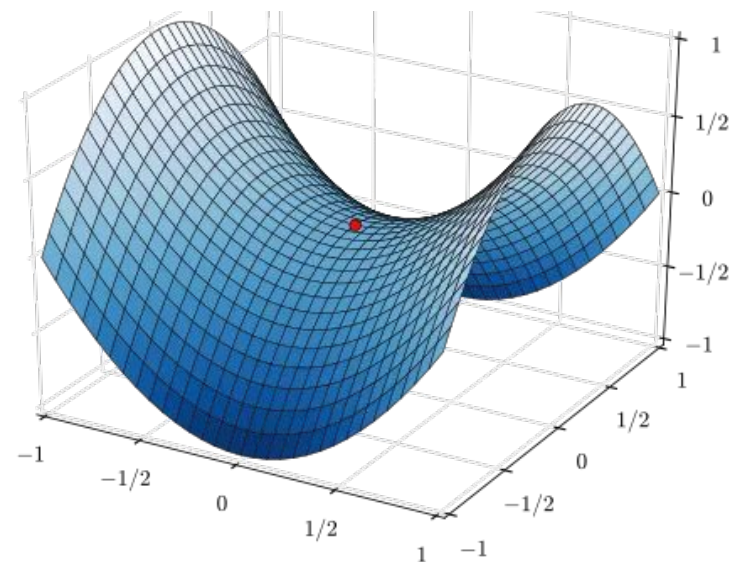
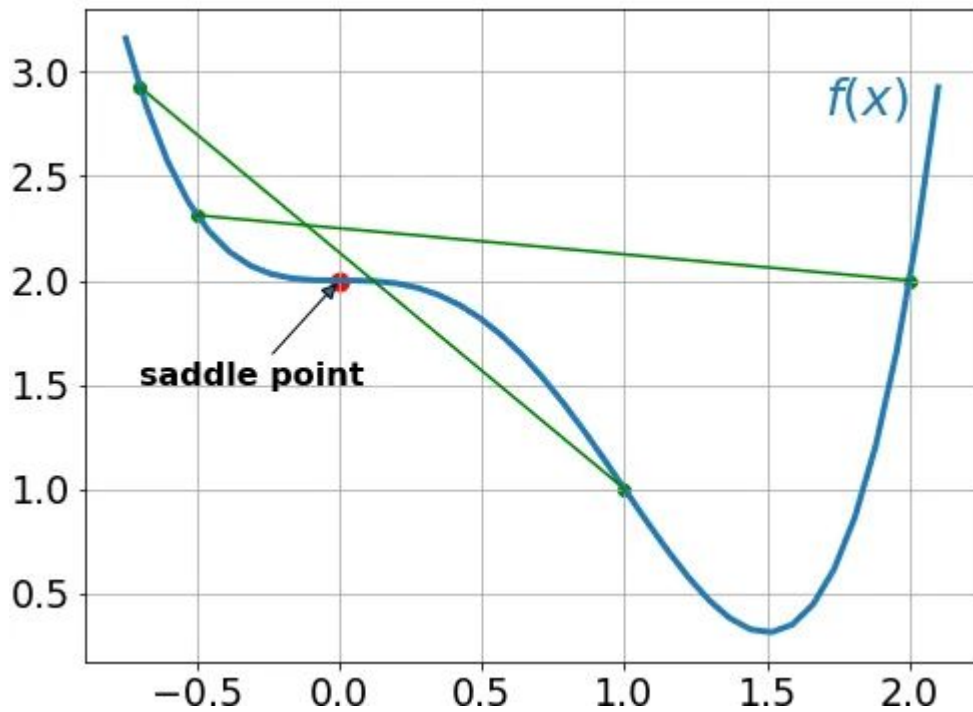


Basics

- Minimize a loss/cost function
- Function Requirements
 1. Differentiable
 2. Convex

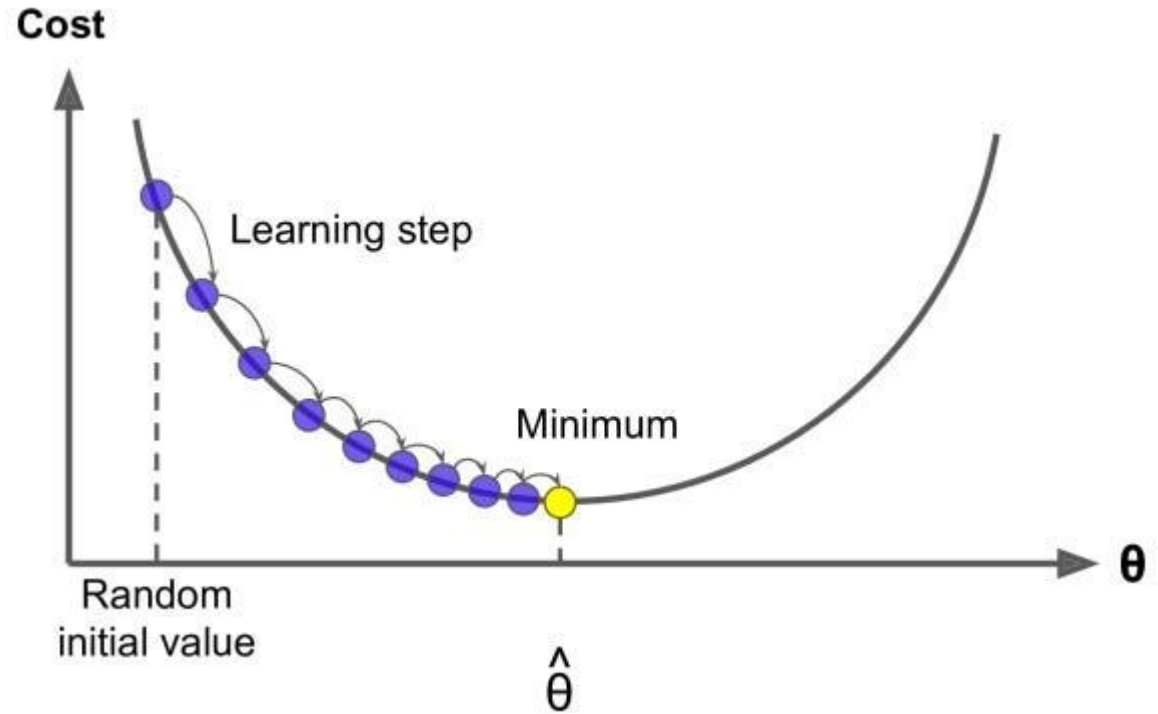


Local Minima and Saddle Points

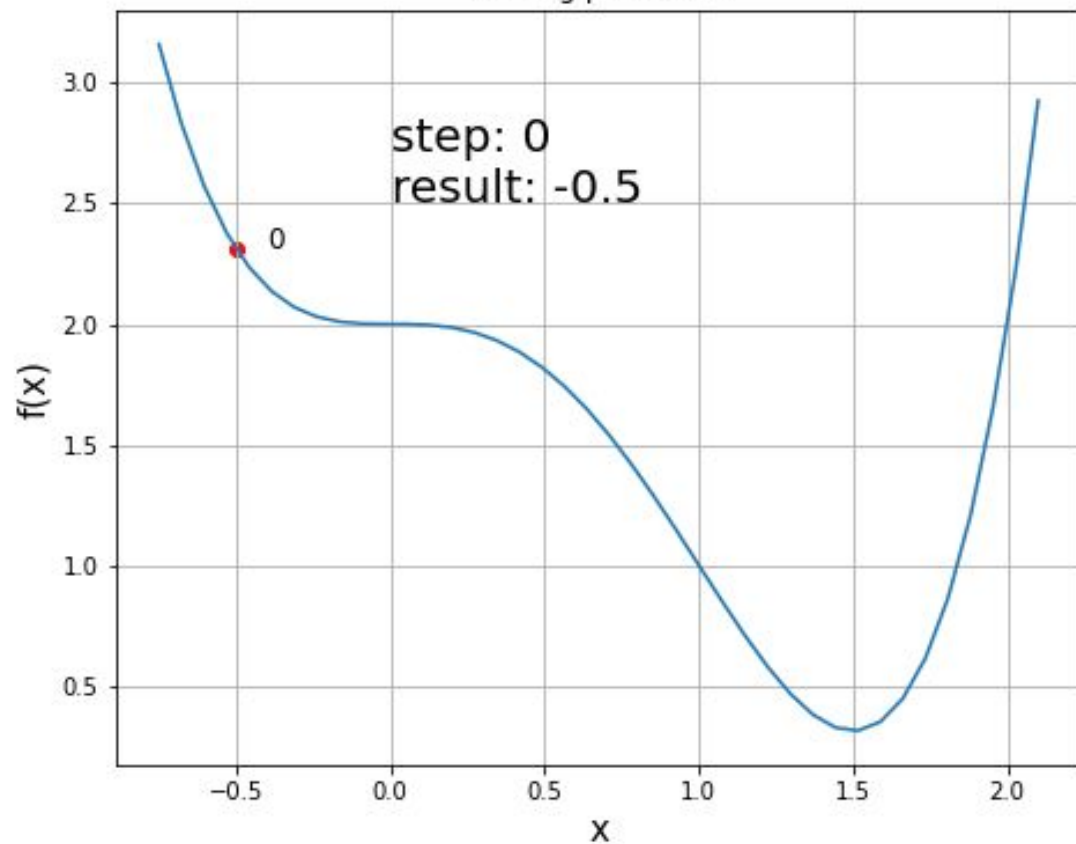


More Basics

- Step-size



Learning rate: 0.4
starting point: -0.5



More Basics

- Step-size
- Partial Derivatives

Scalar-valued multivariable function

$$\nabla f(x_0, y_0, \dots) = \begin{bmatrix} \frac{\partial f}{\partial x}(x_0, y_0, \dots) \\ \frac{\partial f}{\partial y}(x_0, y_0, \dots) \\ \vdots \end{bmatrix}$$

Notation for gradient, called "nabla".

∇f takes the same type of inputs as f

∇f outputs a vector with all possible partial derivatives of f .

More Basics

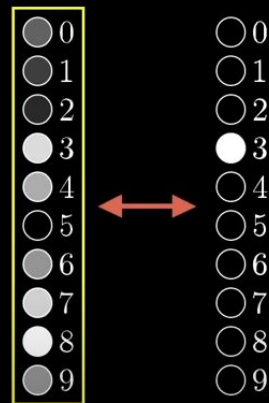
- Step-size
- Gradient
- Partial Derivatives

Cost of 3

3.37

| | | |
|--------|---|---------------------|
| 0.1863 | ← | $(0.43 - 0.00)^2 +$ |
| 0.0809 | ← | $(0.28 - 0.00)^2 +$ |
| 0.0357 | ← | $(0.19 - 0.00)^2 +$ |
| 0.0138 | ← | $(0.88 - 1.00)^2 +$ |
| 0.5242 | ← | $(0.72 - 0.00)^2 +$ |
| 0.0001 | ← | $(0.01 - 0.00)^2 +$ |
| 0.4079 | ← | $(0.64 - 0.00)^2 +$ |
| 0.7388 | ← | $(0.86 - 0.00)^2 +$ |
| 0.9817 | ← | $(0.99 - 0.00)^2 +$ |
| 0.3998 | ← | $(0.63 - 0.00)^2$ |

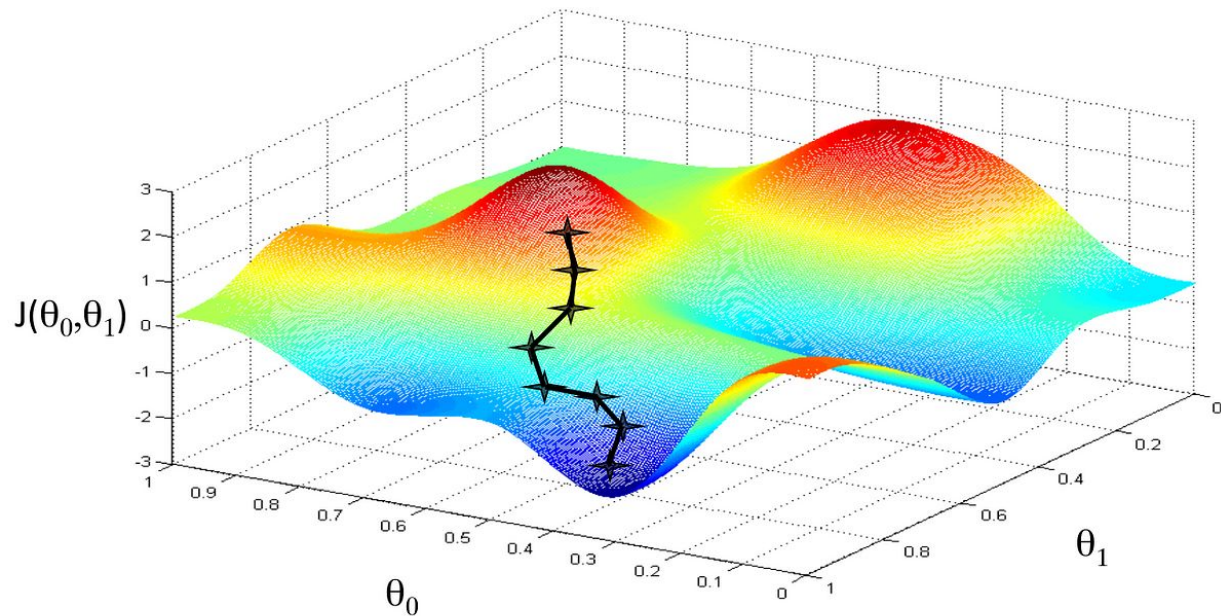
What's the “cost” of this difference?



Utter trash

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/>