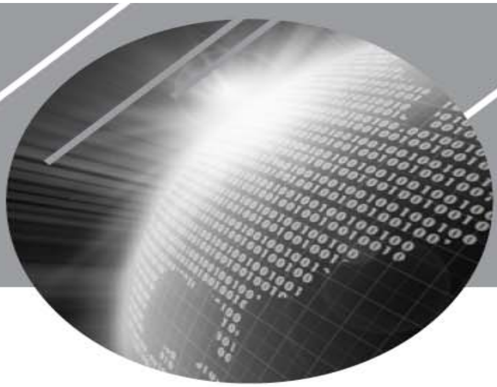
Chapter 7.

Basic Calculus

Part 2. Partial Derivatives and Gradient decent algorithm

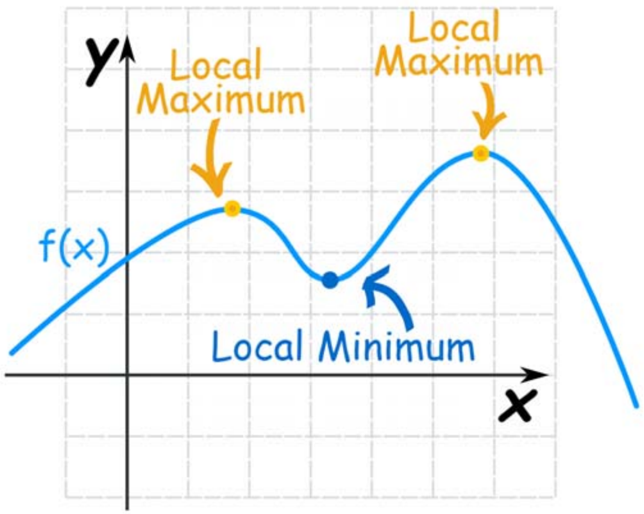


Dept. of Software
Gachon University
Prof. Jaehyuk Choi
Spring 2019

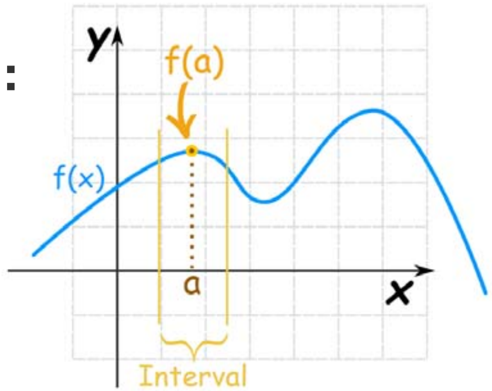
7.3 Finding Minimum algorithmically

Local Maximum and Minimum

- Functions can have "hills and valleys": places where they reach a minimum or maximum value.
- It may not be the minimum or maximum for the **whole function**, but **locally** it is.



local **maximum:**



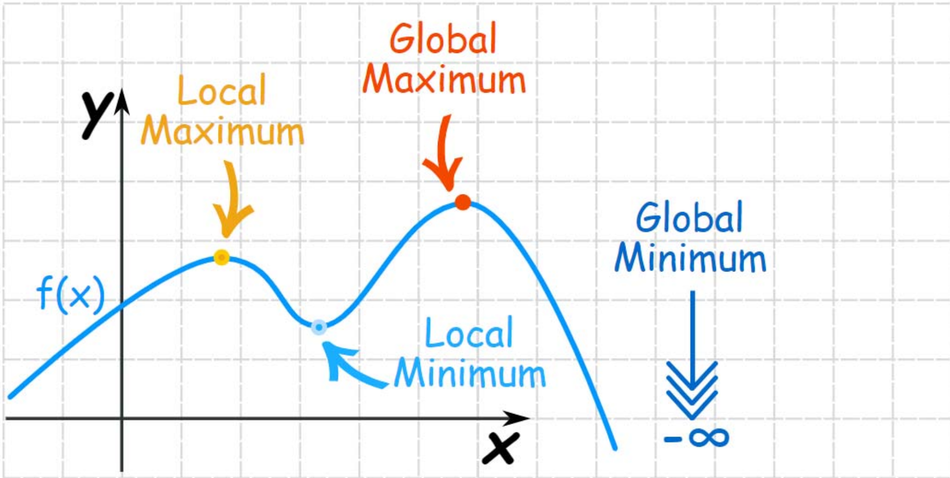
f(a) ≥ f(x) for all x in the interval

Likewise, a local **minimum**

f(a) ≤ f(x) for all x in the interval

Global (or Absolute) Maximum and Minimum

- The maximum or minimum over the **entire data set** is called an "Absolute" or "Global" maximum or minimum.



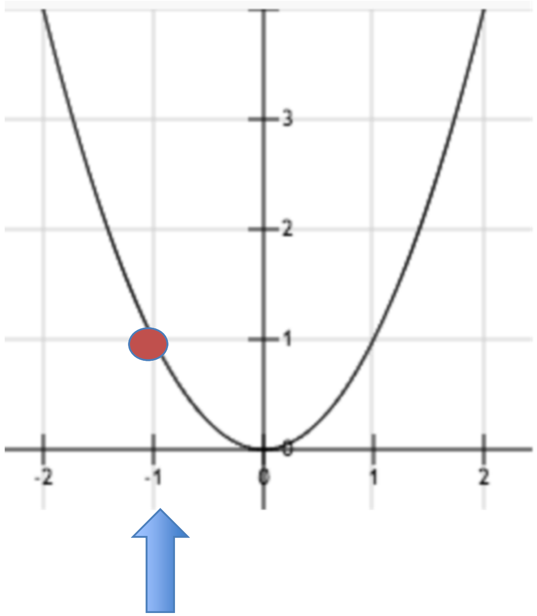
  – There is only one global maximum (and one global minimum) but there can be more than one local maximum or minimum.

Finding Minima/Maxima

- We can find the exact maximum and minimum using calculus (i.e., derivatives)
- The derivative is *zero* at any local maximum or minimum.
- So, one way to find a minimum (or maximum): set $f'(x)=0$ and solve for x.
- **Instead: algorithmically search different values of x until you find one that results in a gradient near 0.**

Finding Minima~~/Maxima~~

- Suppose you start to guess the minimum of given L(x)
  - Assume that finding ∃c s.t. L'(c) = 0 is non-trivial
  - So, you use a parameter θ to some guess; start at the position θ^0 → check $\boldsymbol{L(\theta^0)}$
  - Then, move to θ^1, test $L(\theta^1)$ …



θ^0**=-1**

What is your strategy to choose the next parameter θ^1?

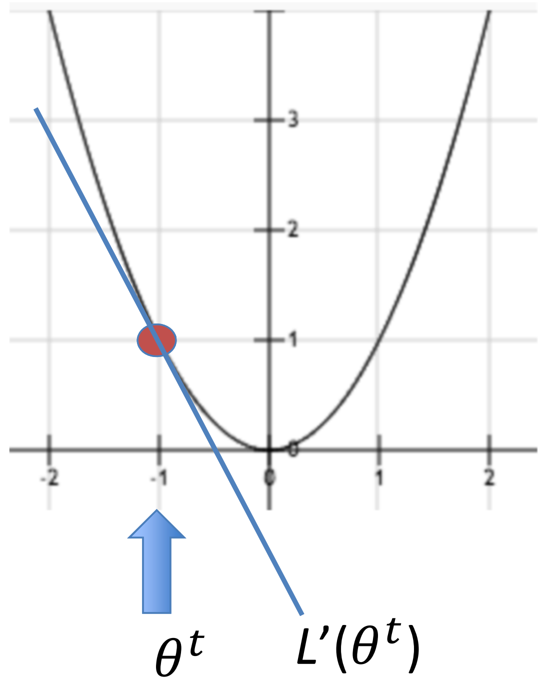
- When solving: $\theta^* = \arg\min_{\theta} L(\theta)$
  - Each time we update the parameters, we obtain θ that makes $L(\theta)$ smaller.

$$L(\theta^0) > L(\theta^1) > L(\theta^2) > \cdots$$

Finding Minima (Cont.)

- If the derivative is positive, the function is **increasing** (as x gets larger).
  - Don't move in that direction, because you'll be moving away from a trough.

- If the derivative is negative, the function is **decreasing.**
  - Keep going, since you're getting closer to a trough



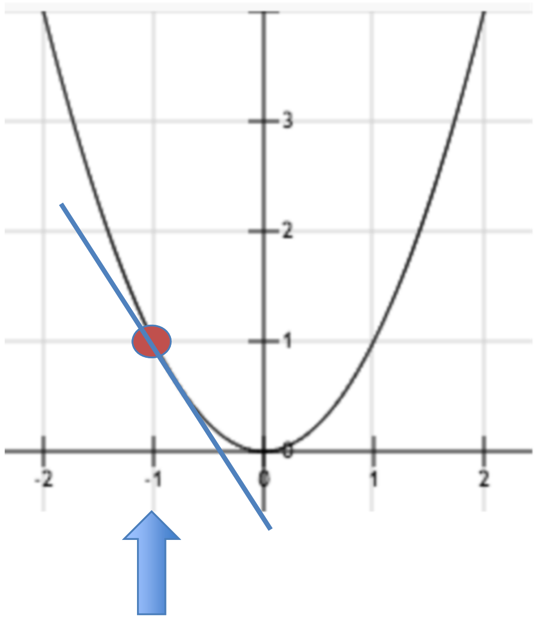
So, it is natural to update the paramter:

$$\theta^{t+1} \leftarrow \theta^t - \eta L'(\theta^t)$$

Update:= 에러를 낮추는 방향 x 한발자국크기 x 현 지점의 기울기
(decent) (step size/rate) (gradient)

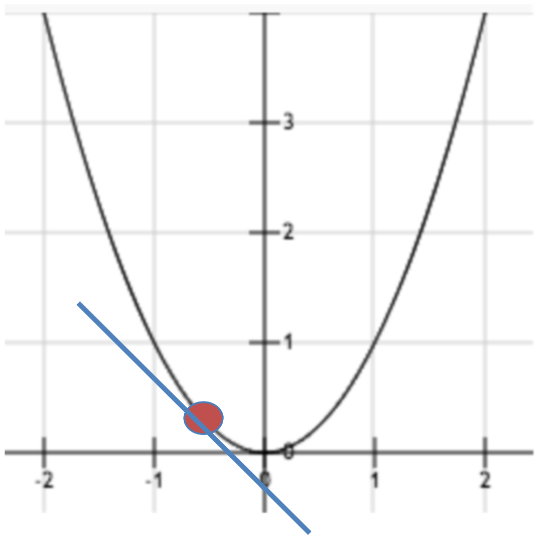
Finding Minima (Cont.)

- Example : Let's use η=0.25



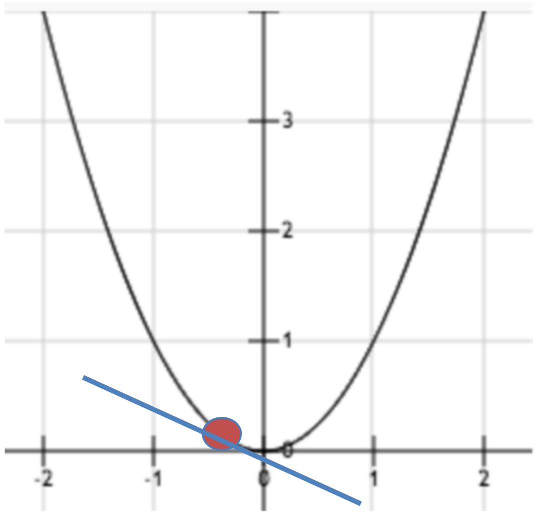
$\theta^0 = -1$

$L'(-1) = -2$



$\theta^1 = -1 - (0.25)(-2) = -0.5$

$L'(-0.5) = -1$



…

$\theta^2 = -0.5 - (0.25)(-1) = -0.125$

….

Eventually, it will reach θ^t =0

Gradient Descent : Outline

1. Initialize the parameters **w** to some guess
   (usually all zeros, or random values)
2. Update the parameters:

$$\mathbf{w} = \mathbf{w} - \eta \, \nabla L(\mathbf{w})$$

3. Update the learning rate (or step-size) η
   (How? Later...: Ch. 7.5)
4. Repeat steps 2-3 until $\nabla L(\mathbf{w})$ is close to zero.

(Cont.)

- Gradient descent is guaranteed to eventually find a *local* minimum if:
  - the learning rate η is decreased appropriately;
  - a finite local minimum exists (i.e., the function doesn't keep decreasing forever).

- Stopping Criteria
  - For most functions, you probably won't get the gradient to be exactly equal to **0** in a reasonable amount of time
  - Once the gradient is sufficiently close to **0**, stop trying to minimize further
  - Stop when the norm of the gradient is below some threshold, ε:

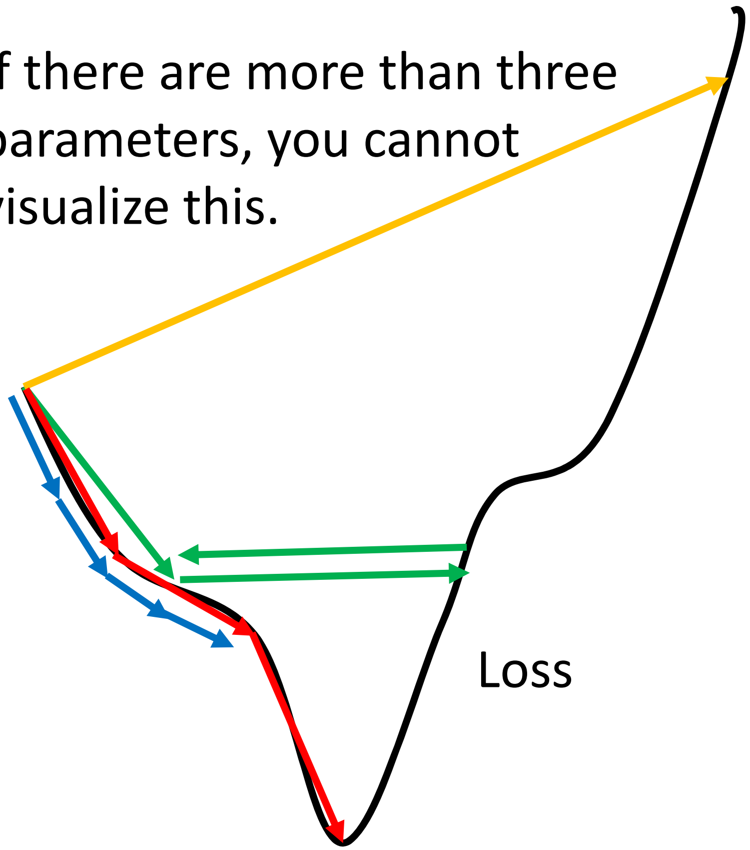
$$\|\nabla L(w)\| < \varepsilon$$

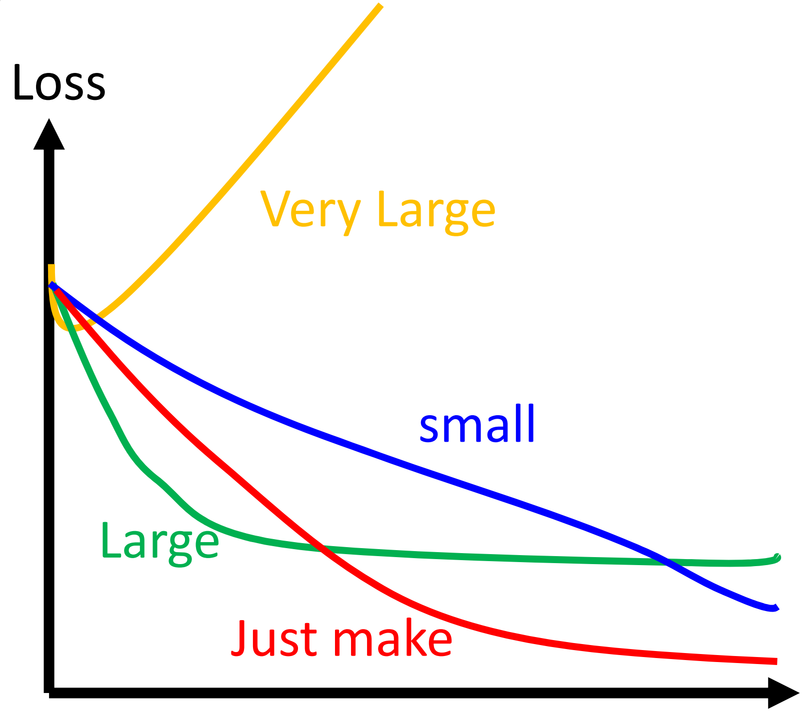
(Appendix) Learning Rate

$$\theta^i = \theta^{i-1} - \eta \nabla L\left(\theta^{i-1}\right)$$

Set the **learning rate η** carefully

If there are more than three parameters, you cannot visualize this.





But you can always visualize this.

7.4 Partial Derivatives (Partial differentiation)

편미분(partial differentiation),
편도함수(paritual derivative)

Introduction

- Multi-Variable Calculus:
  Consider the following functions $f(x_1, x_2, \dots, x_n)$ where $x_1, x_2, \dots, x_n$ are independent variables.

- If we differentiate f with respect variable x_i, then we assume that
  i. $\boldsymbol{x_i}$ as a single variable
  ii. $x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ as constants

https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivative-and-gradient-articles/a/introduction-to-partial-derivatives

Notation

If $f = (x, y)$

- First order partial derivatives:

$$\frac{\partial f}{\partial x} \text{ and } \frac{\partial f}{\partial y}$$

You read the symbol $\frac{\partial f}{\partial x}$ by saying "the partial derivative of *f* with respect to *x*".

Instead of letter d, we use introduce a symbol ∂ to indicate tiny changes.

- Second order partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$$

Example 1.

- Find the first-order partial derivatives for each of the following function:

$$z = 2x^2 + 3xy + 5$$

Solution

- The first partial derivative of *z* w.r.t *x:*

$$z = 2x^2 + 3xy + 5$$

..treat *y* as a constant

$$z = 2x^2 + 3x[y] + 5$$

..then differentiate w.r.t. *x:*

$$\frac{\partial z}{\partial x} = 2(2x) + 3(1)y$$

$$\frac{\partial z}{\partial x} = 4x + 3y$$

Solution

- The first partial derivative of *z* w.r.t *y*:

$$z = 2x^2 + 3xy + 5$$

..treat *x* as a constant

$$z = 2[x^2] + 3[x]y + 5$$

..then differentiate *z* w.r.t. *y:*

$$\frac{\partial z}{\partial y} = 0 + 3x(1) + 0$$

$$\frac{\partial z}{\partial y} = 3x$$

Example 2.

Write down all first-order partial derivatives $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$

and second-order partial derivatives of the following function:

$$f(x,y) = x^3 y^3 + x \sin y + y \ln x$$

Example

$$f(x,y) = x^3y^3 + x\sin y + y\ln x$$

Solution

First order PD:

$$\frac{\partial f}{\partial x} = 3x^2y^3 + \sin y + \frac{y}{x}$$

$$\frac{\partial f}{\partial y} = 3x^3y^2 + x\cos y + \ln x$$

Solution

Second order PD:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$$

$$= \frac{\partial}{\partial x}\left(3x^2y^3 + \sin y + \frac{y}{x}\right)$$

$$= 6xy^3 - \frac{y}{x^2}$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$$

$$= \frac{\partial}{\partial y}\left(3x^3y^2 + x\cos y + \ln x\right)$$

$$= 6x^3y - x\sin y$$

Solution

Second order PD:

(mixed partial)

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$$

$$= \frac{\partial}{\partial x}\left(3x^3 y^2 + x\cos y + \ln x\right)$$

$$= 9x^2 y^2 + \cos y + \frac{1}{x}$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$$

$$= \frac{\partial}{\partial y}\left(3x^2 y^3 + \sin y + \frac{y}{x}\right)$$

$$= 9x^2 y^2 + \cos y + \frac{1}{x}$$

Note:

In the previous example,
we observed that $$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

This properties hold for all functions provided that certain smoothness properties are satisfies.

The mixed partial derivative must be equal whenever *f* is continuous.

(i) Solution

$$(i) \quad f(x,y) = 5x^2e^{4y} + 3xy - 2x^3$$

$$\frac{\partial f}{\partial x} = 10xe^{4y} + 3y - 6x^2 \qquad \frac{\partial f}{\partial y} = 20x^2e^{4y} + 3x$$

$$\frac{\partial^2 f}{\partial x^2} = 10e^{4y} - 12x \qquad \frac{\partial^2 f}{\partial y^2} = 80x^2e^{4y}$$

$$\frac{\partial^2 f}{\partial x \partial y} = 40xe^{4y} + 3 \qquad \frac{\partial^2 f}{\partial y \partial x} = 40xe^{4y} + 3$$

(ii) Solution

$$(ii) \quad f(x,y) = xe^{xy} + x\sin(xy) + x^2y$$

$$\frac{\partial f}{\partial x} = xye^{xy} + e^{xy} + xy\cos(xy) + \sin(xy) + 2xy$$

$$\frac{\partial f}{\partial y} = x^2e^{xy} + x^2\cos(xy) + x^2$$

$$\frac{\partial^2 f}{\partial x^2} = xy^2e^{xy} + 2ye^{xy} - xy^2\sin(xy) + 2y\cos(xy) + 2y$$

$$\frac{\partial^2 f}{\partial y^2} = x^3e^{xy} - x^3\sin(xy)$$

$$\frac{\partial^2 f}{\partial x \partial y} = x^2ye^{xy} + 2xe^{xy} - x^2y\sin(xy) + 2x\cos(xy) + 2x$$

$$\frac{\partial^2 f}{\partial y \partial x} = x^2ye^{xy} + 2xe^{xy} - x^2y\sin(xy) + 2x\cos(xy) + 2x$$

References

- Lecture Notes from Greg Kelly’s Lecture Notes, Richland, Washington
- http://speech.ee.ntu.edu.tw/~tlkagk/courses_ML17.html
- Dr. Nazuhusna and Dr. Shahrir Rizal’s Lecture Notes (EQT 101 Engineering Mathematics I) @ UniMAP: Universiti Malaysia Perlis
- and
- KhanAcademy