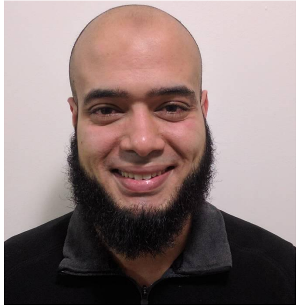
Machine Learning

Calculus

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)



Compute Partial Derivatives

- $\partial/\partial \mathbf{x}\quad 4x^2$
- $\partial/\partial \mathbf{x}\quad 4(3x + 1)^3$
- $\partial/\partial \mathbf{x}\quad 4x^2 + 2(3 - y)^2$
- $\partial/\partial \mathbf{y}\quad 4x^2 + 2(3 - y)^2$
- $\partial/\partial \mathbf{x}\quad 2x^5 - 4xyz$

- $8x$
- $36 * (3x + 1)^2$
- $8x$
- $-4(3 - y)$
- $10x^4 - 4yz$

Compute Partial Derivatives

- $\partial/\partial \mathbf{x} \quad \sin(x)$
- $\partial/\partial \mathbf{x} \quad \sin(x^5)$
- $\partial/\partial \mathbf{x} \quad \cos(\sin(x^5))$

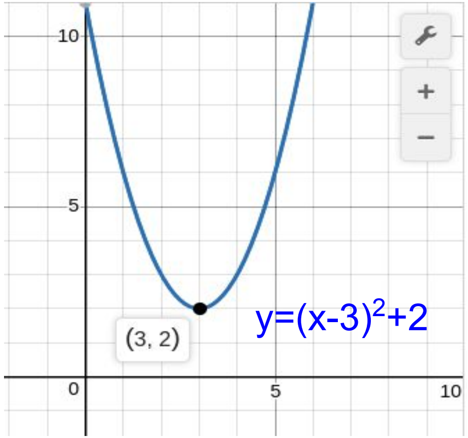
- $\cos(x)$
- $5x^4(\cos(x^5))$
- $-\cos(x^5)\ \sin(\ \sin(x^5)\ )\ 5x^4$

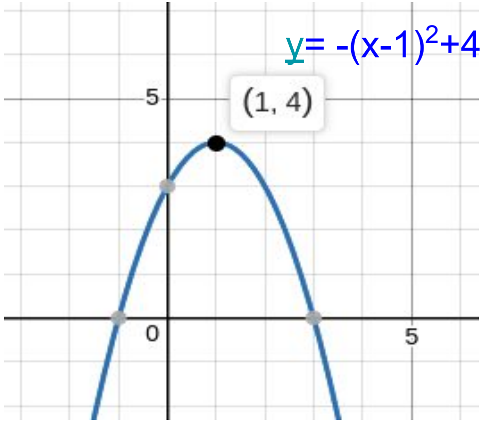
Question!

- Your mother is preparing **pizza**
- She wants to add the right (optimal) amount of salt to it
  - Too little salt will make the pizza very bland (دلعة)
  - Too much salt will make the pizza salty (مملحة)
- Let’s model this as an equation
  - X: input, the amount of salt
  - Y: error criteria: to what degree not optimal level of salty
  - Goal: find the optimal amount of salt that makes the best-tasting pizza
- Which function represents the phenomena?
  - Linear? Quadratic? Cubic? Quartic?

Quadratic function

- The **minimum** of the first function is at y = 2
- This is known as **global minimum** value
    - You can't find a value with smaller y
- The graph of the quadratic function is decreasing on one side of the axis and increasing on the other side (**U-shaped**) - **convex function**
    - ML focuses on the decrease/increase case (**minimization**)



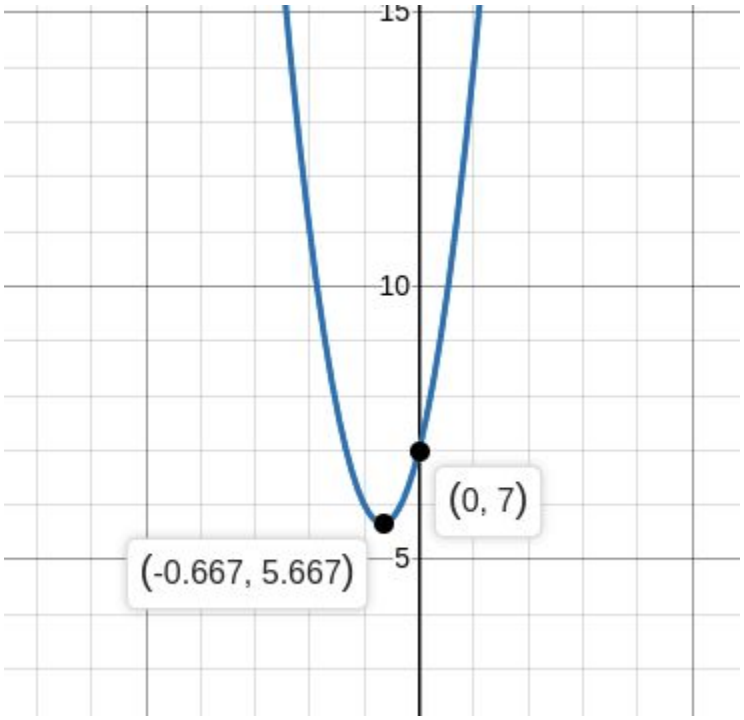


Find min/max of quadratic function

- To find the minimum (or max) point, we need to compute the **derivative**
  - The **derivative** is what measures the **steepness** of the graph of a **function** at a **specific point** on the graph (so it is a **slope = rate of change**)
- Assume we have a function: $\mathbf{f(x) = 3x^2 + 4x + 7}$
- What is its derivative?
- $\mathbf{f'(x) = 6x + 4}$
  - $aX^b = (a * b)X^{b-1}$
- Set to zero and solve: $6x + 4 = 0$
  - $x = -2/3 = -0.6666666$ (the minimum)
  - This is an **analytical** solution (*steps to give the exact solution*)
- Minimum or maximum?
  - We can work that out by double differentiation

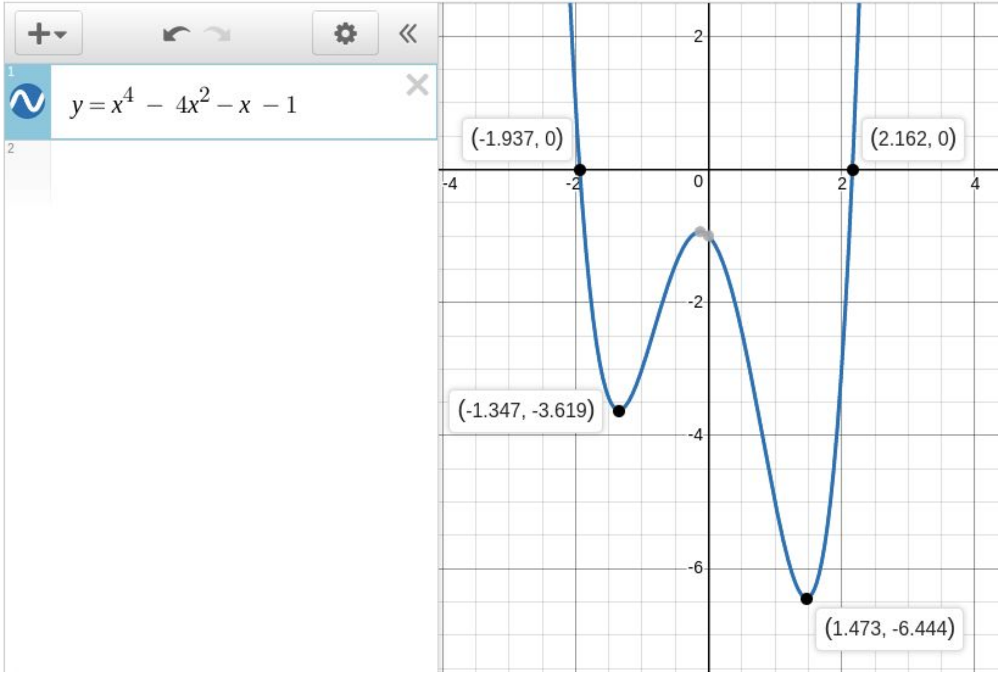
Tangent slopes

- $f(x) = 3x^2 + 4x + 7$
- $f'(x) = 6x + 4$
  - The minimum: $x = -2/3 = -0.6666666$
  - We can use it to know the sign at specific points
- Evaluate at f'
  - $x = -2/3 \Rightarrow 0$ (**zero** slope)
  - $x = 0 \Rightarrow 4$ (**positive** slope)
  - $x = -2 \Rightarrow -8$ (**negative** slope)



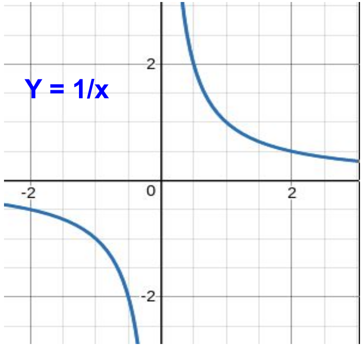
Visualize functions

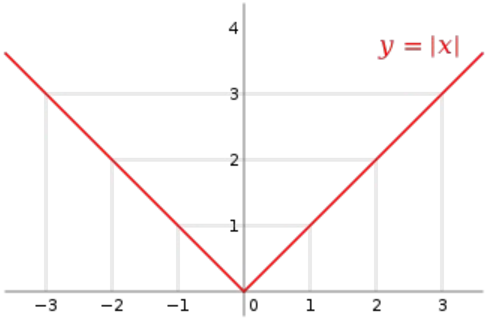
- You might use some online sites to visualize functions
  - For example: https://www.transum.org/Maths/Activity/Graph/Desmos.asp



Differentiable Function

- A **differentiable** function of one real variable is a function whose **derivative exists** at each point in its domain
- 3 cases for non-differentiable functions
    - discontinuity, corner, crusp, tangent line has vertical slope





The **absolute value** function is **continuous** (i.e. it has no gaps). It is differentiable everywhere except at the point x = 0

Derivative Rules

1. Constant Rule: $\dfrac{d}{dx}(c) = 0$, where c is a constant

2. Power Rule: $\dfrac{d}{dx}(x^n) = nx^{n-1}$

3. Product Rule: $(fg)' = f'g + fg'$

4. Quotient Rule: $\left(\dfrac{f}{g}\right)' = \dfrac{f'g - fg'}{g^2}$

5. Chain Rule: $(f(g(x)))' = f'(g(x))g'(x)$

Common Derivatives

Polynomials

$\frac{d}{dx}(c)=0$ $\frac{d}{dx}(x)=1$ $\frac{d}{dx}(cx)=c$ $\frac{d}{dx}(x^n)=nx^{n-1}$ $\frac{d}{dx}(cx^n)=ncx^{n-1}$

Exponential/Logarithm Functions

$\frac{d}{dx}(a^x)=a^x\ln(a)$ $\frac{d}{dx}(\mathbf{e}^x)=\mathbf{e}^x$

$\frac{d}{dx}(\ln(x))=\frac{1}{x},\ x>0$ $\frac{d}{dx}(\ln|x|)=\frac{1}{x},\ x\neq 0$ $\frac{d}{dx}(\log_a(x))=\frac{1}{x\ln a},\ x>0$

Function Composition

- An operation in which two functions, (in this case, f and g), combine to generate a new function, (h), such that:
  $h(x) = g(f(x))$.
- This means that function g is applied to **the output of** function f for x
- Example: y = sin(sigmoid(**sqrt(x)**))
  - Given x:
  - Compute s = sqrt(x)
  - Then compute t = sigmoid(s)
  - Then compute y = sin(t)

Chain Rule

- A rule that makes our life easy when we compute the derivative of a composition of functions
- Example:
  - Let y = sin(sigmoid(**sqrt(x)**))
  - Compute ∂y/∂x
- Rule
  - $$\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)$$

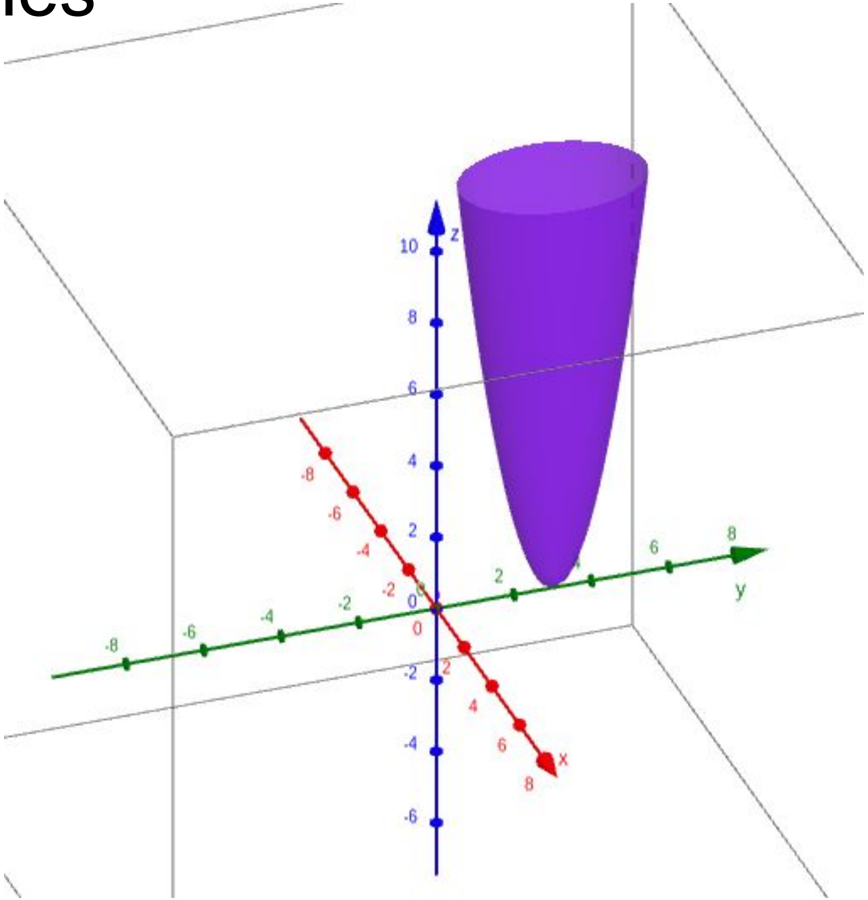
$$\frac{d}{dx}\left[f\left(g(h(x))\right)\right] = f'\left(g(h(x))\right)g'(h(x))h'(x)$$

Example

- Compute $\partial y/\partial x$ where $y = (3x+5)^4$
- Use series of **symbols and compute partial derivatives relative to them then multiply their results**
  - **$y = a^4$**
  - **$a = 3x+5$**
- Compute $\partial y/\partial x = \partial y/\partial a * \partial a/\partial x$
- $\partial y/\partial a = 4a^3$
- $\partial a/\partial x = 3$
- $\partial y/\partial x = 4a^3 * 3 = 4(3x+5)^3 * 3 = 12\,(3x+5)^3$

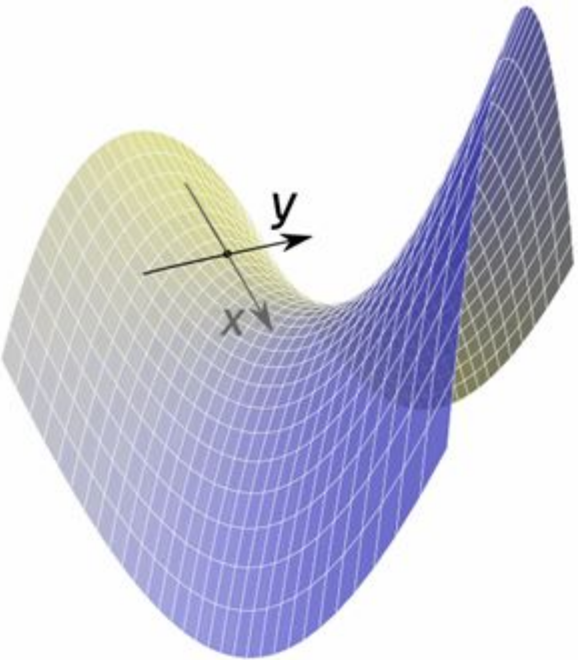
Function of Several Variables

- Sometimes, our function consists of several variables
- Assume our function is:
  $f(x, y) = z = 4x^2 + 2(y - 3)^2$
  - 2 independent variables (x, y)
  - Dependent variable: z
- Let’s draw using an online site



Partial Derivatives

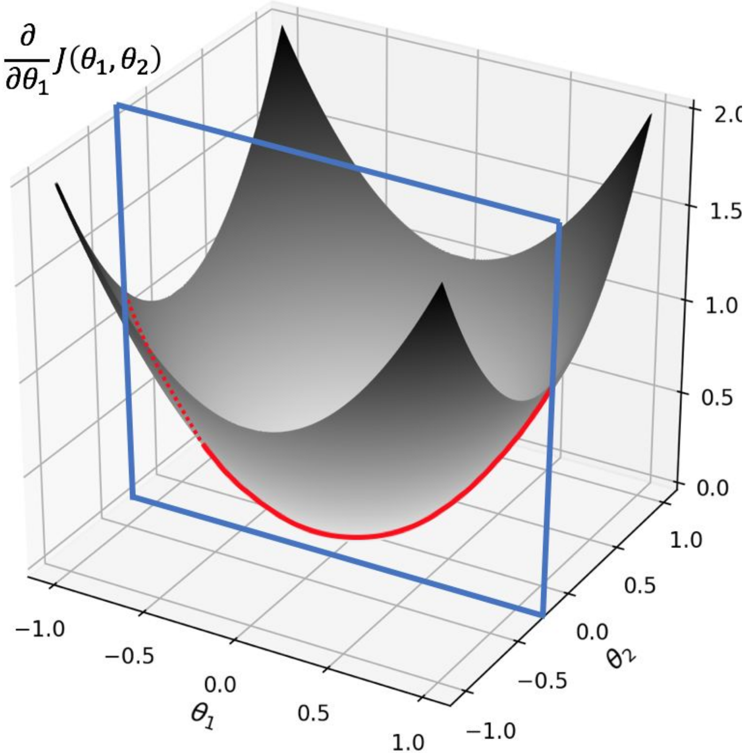
- Partial Derivative: derivatives of a **multivariate** function with respect to one of its arguments (*other variables are constants*)
  - When we find the slope in the **x direction** (while keeping y fixed) we have found a partial derivative
    - The opposite for y
  - Derivatives are called the **partial** derivative
- Syn
  - $f_x = \frac{\partial f}{\partial x}$ $\quad f_y = \frac{\partial f}{\partial y}$



Img src

Partial Derivatives: Intuition

- We have 2 variables: θ1 and θ2
- When we take partial derivatives for θ1, we fix a given θ2
- Intuitively, we took a slice of the 2D Graph (a 1D slice)
- Now we apply normal univariate derivatives



Img src

Partial Derivatives

- Can you compute the partial derivative of $f(x, y) = 4x^2 + 2(y - 3)^2$
  - Once for x and once y

$$\frac{\partial}{\partial x}\left[4x^2 + 2(y-3)^2\right]$$

$$\frac{\partial}{\partial y}\left[4x^2 + 2(y-3)^2\right]$$

Differentiate the sum term by term and factor out constants:

$$= 4\left(\frac{\partial}{\partial x}(x^2)\right) + \frac{\partial}{\partial x}\left(2(-3+y)^2\right)$$

Use the power rule, $\frac{\partial}{\partial x}(x^n) = n\,x^{n-1}$, where $n = 2$.

$\frac{\partial}{\partial x}(x^2) = 2x$:

$$= \frac{\partial}{\partial x}\left(2(-3+y)^2\right) + 4\,\boxed{2x}$$

Simplify the expression:

$$= 8x + \frac{\partial}{\partial x}\left(2(-3+y)^2\right)$$

The derivative of $2(y-3)^2$ is zero:

$$= 8x + \boxed{0}$$

Let's use an [online calculator]

Differentiate the sum term by term and factor out constants:

$$= \frac{\partial}{\partial y}(4x^2) + 2\left(\frac{\partial}{\partial y}((-3+y)^2)\right)$$

The derivative of $4x^2$ is zero:

$$= 2\left(\frac{\partial}{\partial y}((-3+y)^2)\right) + \boxed{0}$$

Simplify the expression:

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

Using the chain rule, $\frac{\partial}{\partial y}((y-3)^2) = \frac{\partial u^2}{\partial u}\frac{\partial u}{\partial y}$, where $u = y-3$ and $\frac{\partial}{\partial u}(u^2) = 2u$:

$$= 2\boxed{\left(2(-3+y)\left(\frac{\partial}{\partial y}(-3+y)\right)\right)}$$

Simplify the expression:

$$= 4(-3+y)\left(\frac{\partial}{\partial y}(-3+y)\right)$$

Differentiate the sum term by term:

$$= \boxed{\frac{\partial}{\partial y}(-3) + \frac{\partial}{\partial y}(y)}\,4(-3+y)$$

The derivative of -3 is zero:

$$= 4(-3+y)\left(\frac{\partial}{\partial y}(y) + \boxed{0}\right)$$

Simplify the expression:

$$= 4(-3+y)\left(\frac{\partial}{\partial y}(y)\right)$$

The derivative of y is 1:

$$= \boxed{1}\,4(-3+y)$$

Homework

- Compute the partial derivatives of these functions and compare with the tool
- $(2x-4)^5 + 4yx$
- $6(2x^3-4)^5 + 4yx$
- sqrt(4yx)
- $\log(2x^3 + 4yx)$
- $\exp(2x^3 + 4yx)$

Relevant Resources

- intuition about derivatives - More - By Andrew NG
- Chain Rule- StatQuest

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."