Machine Learning Foundations
Calc II: Partial Derivatives & Integrals

Using Gradients in
Python to Enable Algorithms to
Learn from Data

Jon Krohn, Ph.D.

jonkrohn.com/talks

github.com/jonkrohn/ML-foundations

Machine Learning Foundations
Calc II: Partial Derivatives & Integrals

Slides: `jonkrohn.com/talks`

Code: `github.com/jonkrohn/ML-foundations`

Stay in Touch:

`jonkrohn.com` to sign up for email newsletter

`linkedin.com/in/jonkrohn`

`jonkrohn.com/youtube`

`twitter.com/JonKrohnLearns`

The Pomodoro Technique

Rounds of:

- 25 minutes of work
- with 5 minute breaks

Questions best handled at breaks, so save questions until then.

When people ask questions that have already been answered, do me a favor and let them know, politely providing response if appropriate.

Except during breaks, I recommend attending to this lecture only as topics are not discrete: Later material builds on earlier material.

POLL

Where are you?

- The Americas
- Europe / Middle East / Africa
- Asia-Pacific
- Extra-Terrestrial Space

POLL

What are you?

- Developer / Engineer
- Scientist / Analyst / Statistician / Mathematician
- Combination of the Above
- Other

POLL

What is your level of familiarity with Calculus?

- Little to no exposure
- Some understanding of the theory
- Deep understanding of the theory
- Deep understanding of the theory and experience applying calculus operations (e.g., differentiation) with code

POLL

What is your level of familiarity with Machine Learning?

- Little to no exposure, or exposure to theory only
- Experience applying machine learning with code
- Experience applying machine learning with code and some understanding of the underlying theory
- Experience applying machine learning with code and strong understanding of the underlying theory

ML Foundations Series

Calculus II **builds upon** and is **foundational for**:

1. Intro to Linear Algebra
2. Linear Algebra II: Matrix Operations
3. **Calculus I: Limits & Derivatives**
4. **Calculus II: Partial Derivatives & Integrals**
5. Probability & Information Theory
6. Intro to Statistics
7. Algorithms & Data Structures
8. **Optimization**

Calc II: Partial Derivatives & Integrals

1. Review of Introductory Calculus
2. Machine Learning Gradients
3. Integrals

Calc II: Partial Derivatives & Integrals

1. **Review of Introductory Calculus**
2. Machine Learning Gradients
3. Integrals

Segment 1: Review of Introductory Calc

- The Delta Method
- Differentiation with Rules
- AutoDiff: Automatic Differentiation

What Calculus Is

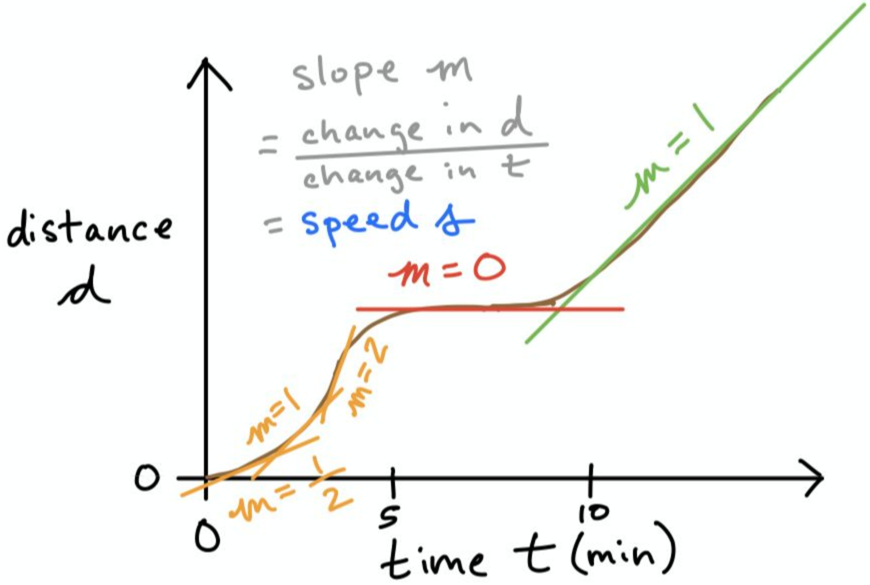
- **Mathematical study of continuous change**
- Two branches:
  a. **Differential calculus**: expanded on in *Calc II*
  b. **Integral calculus**: a focus of *Calc II* subject

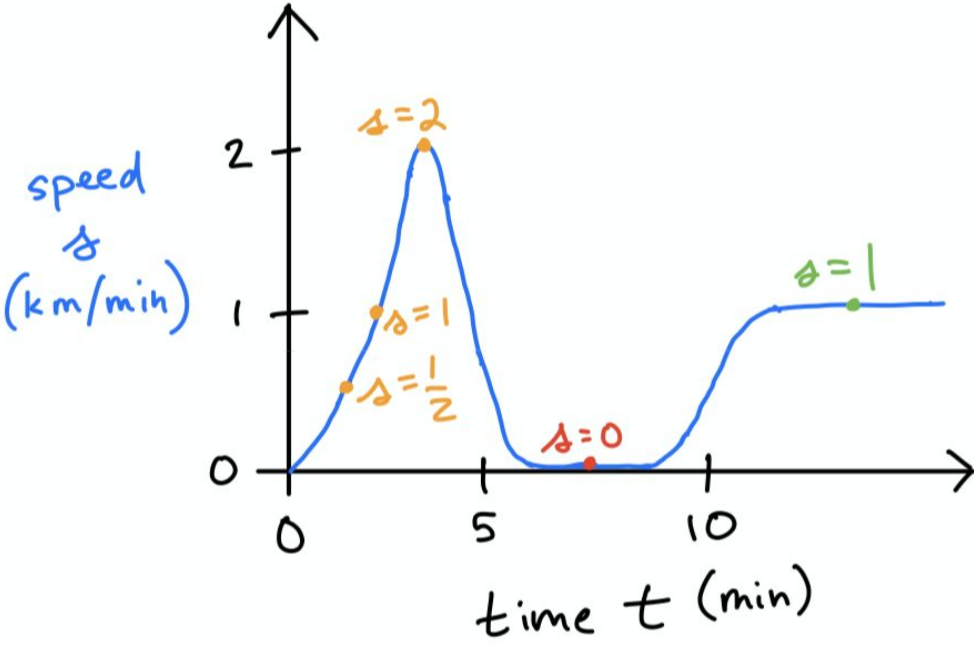
What Calculus Is

- **Mathematical study of continuous change**
- Two branches:
    a. **Differential calculus**: expanded on in *Calc II*
    b. **Integral calculus**: a focus of *Calc II* subject

What Differential Calculus Is

- **Study of rates of change**
- Consider a vehicle traveling some distance *d* over time *t*:



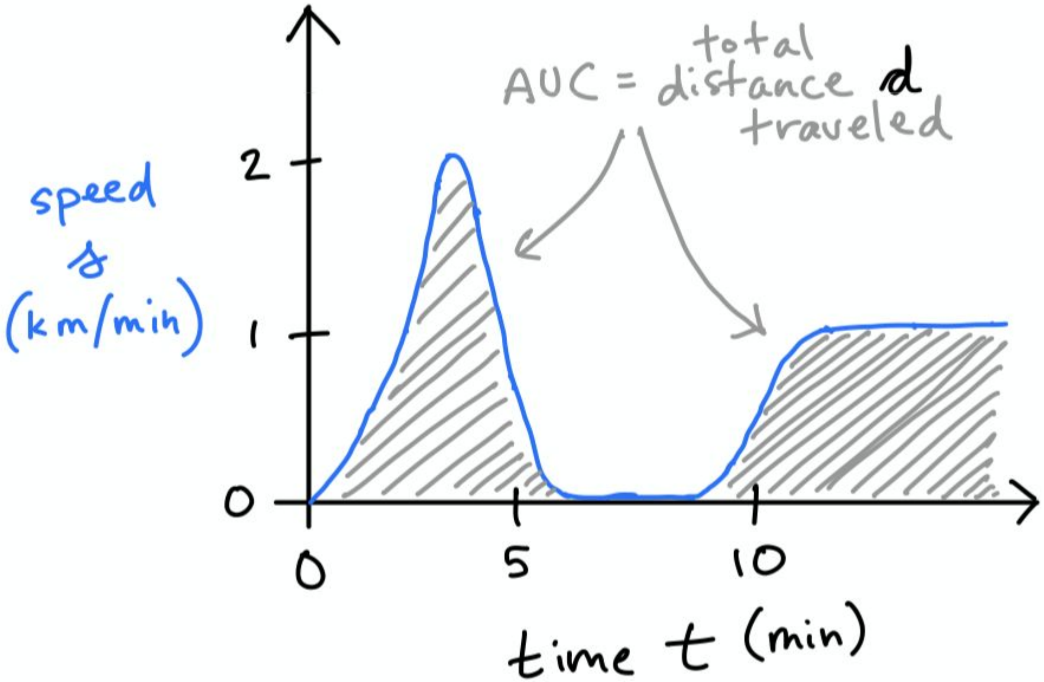


What Calculus Is

- **Mathematical study of continuous change**
- Two branches:
  a. **Differential calculus**: expanded on in *Calc II*
  b. **Integral calculus**: a focus of *Calc II* subject

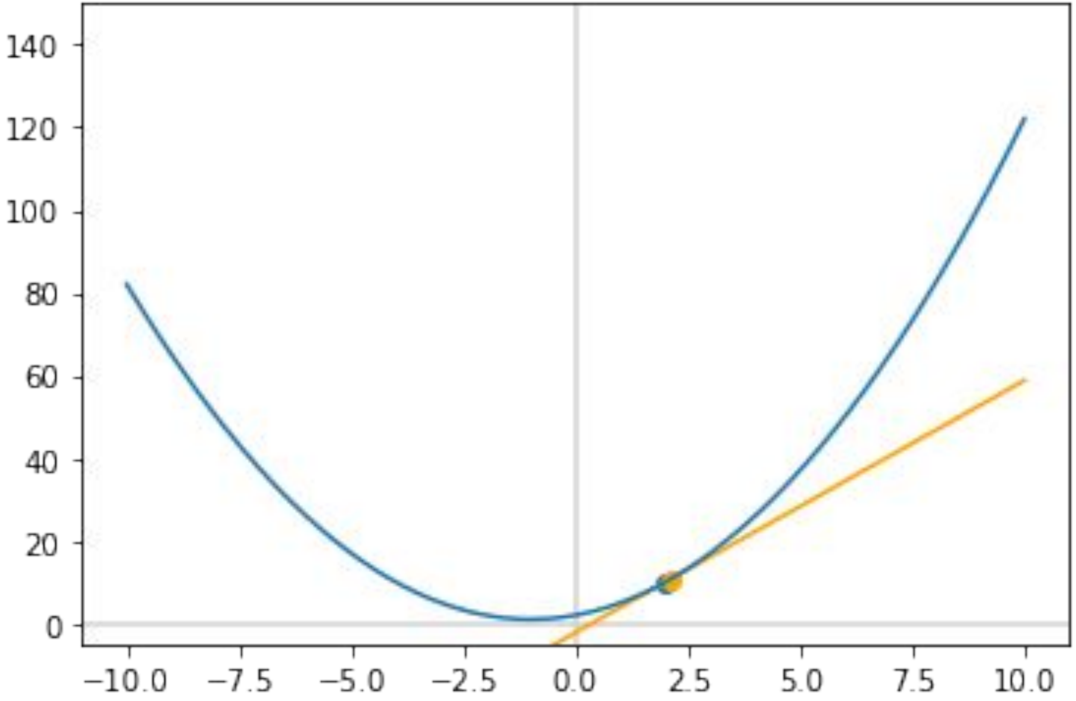
What Integral Calculus Is

- **Study of areas under curves**
- Facilitates the “opposite” of differential calculus:



The Delta Method

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



Derivative Notation

$$y = f(x)$$

differentiation operator:					$\frac{d}{dx}$	D_x
first derivative of y w.r.t. x:	$\dot{y}$	y'	$f'(x)$	$\frac{dy}{dx}$	$\frac{d}{dx}f(x)$	$D_x f$
second derivative of y w.r.t. x:	$\ddot{y}$	y''	$f''(x)$	$\frac{d^2y}{dx^2}$	$\frac{d^2}{dx^2}f(x)$	$D_x^2 f$

Derivative of a Constant

Assuming *c* is constant: $$\frac{d}{dx}c = 0$$

Intuition: A constant has no variation so its slope is nothing, e.g.:

$$\frac{d}{dx}25 = 0$$

The Power Rule

$$\frac{d}{dx} x^n = n x^{n-1}$$

E.g.: $$\frac{d}{dx} x^4 = 4x^{4-1} = 4x^3$$

The Constant Product Rule

$$y = 2x^4$$

$$\frac{dy}{dx} = 2(4x^3) = 8x^3$$

The Sum Rule

$$\frac{d(y+w)}{dx} = \frac{dy}{dx} + \frac{dw}{dx}$$

E.g.: $y = x^4$ $\quad w = x^9$

$$\frac{d(y+w)}{dx} = \frac{dy}{dx} + \frac{dw}{dx}$$

$$= \frac{d}{dx}(x^4) + \frac{d}{dx}(x^9)$$

$$= 4x^3 + 9x^8$$

The Chain Rule

- Many applications within ML
  - Critical for backpropagation algo used to train neural nets
- Based on *nested functions*:
  - Let's say $y = (5x + 25)^3$
  - We can let $u = 5x + 25$
  - In that case, $y = u^3$
  - y is a function of u, and u is a function of x
- *Chain rule* is easy way to find derivative of nested function:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

The Chain Rule

$$y = (2x^2 + 8)^2$$

$$y = u^2$$

$$\frac{dy}{du} = 2u$$

$$= 2(2x^2 + 8)$$

$$= 4x^2 + 16$$

$$u = 2x^2 + 8$$

$$\frac{du}{dx} = 4x$$

$$\frac{dy}{dx} = \frac{dy}{du}\,\frac{du}{dx}$$

$$= (4x^2 + 16)(4x)$$

$$= 16x^3 + 64x$$

Power Rule on a Function Chain

$$\frac{d}{dx}u^n = nu^{n-1}\frac{du}{dx}$$

E.g.: $y = (3x+1)^2$

$n = 2 \qquad u = 3x+1$

$$\frac{du}{dx} = 3$$

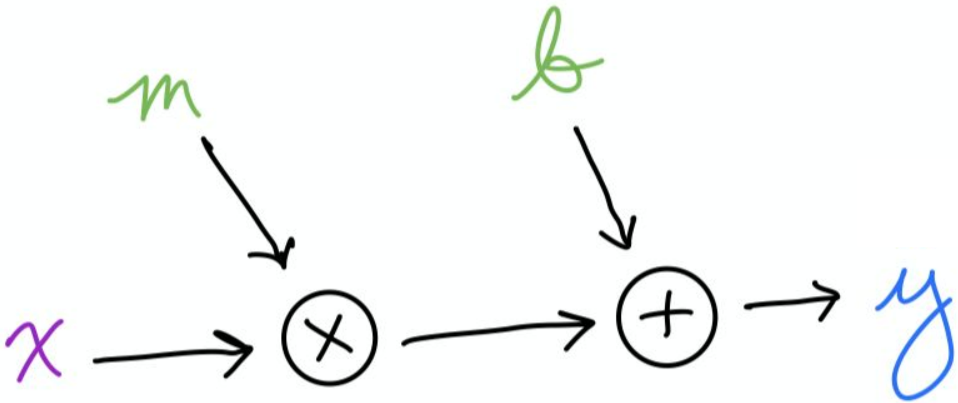
$$\frac{dy}{dx} = \frac{d}{dx}u^n = nu^{n-1}\frac{du}{dx}$$

$$= \left(2(3x+1)^1\right)(3)$$

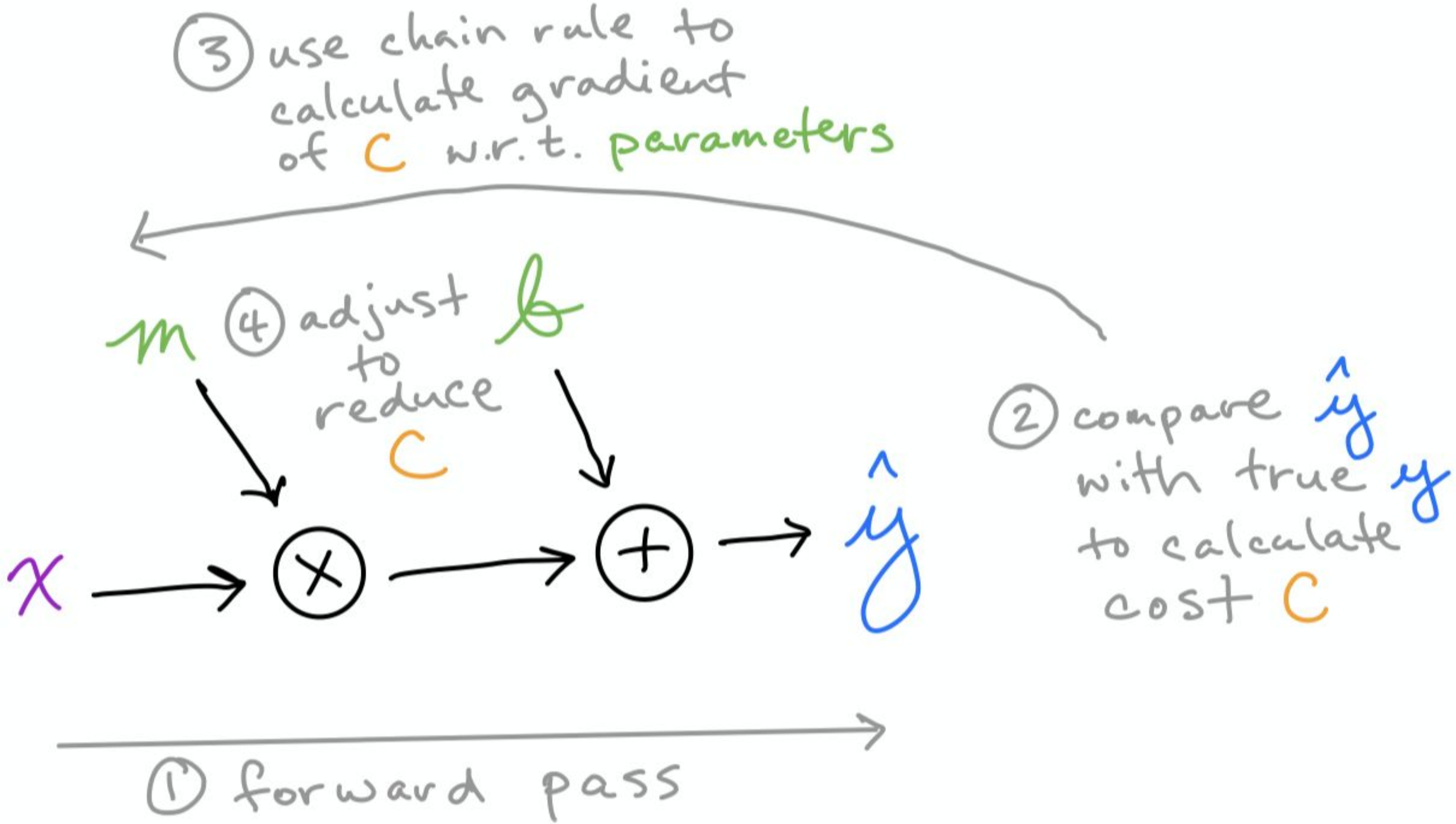
$$= (6x+2)(3) = 18x+6$$

Fitting a Line with Machine Learning

- Line equation $y = mx + b$ as *directed acyclic graph* (DAG)
- Nodes are input, output, parameters, or operations
- Directed edges ("arrows") are tensors (N.B.: non-operation nodes can be tensors too)



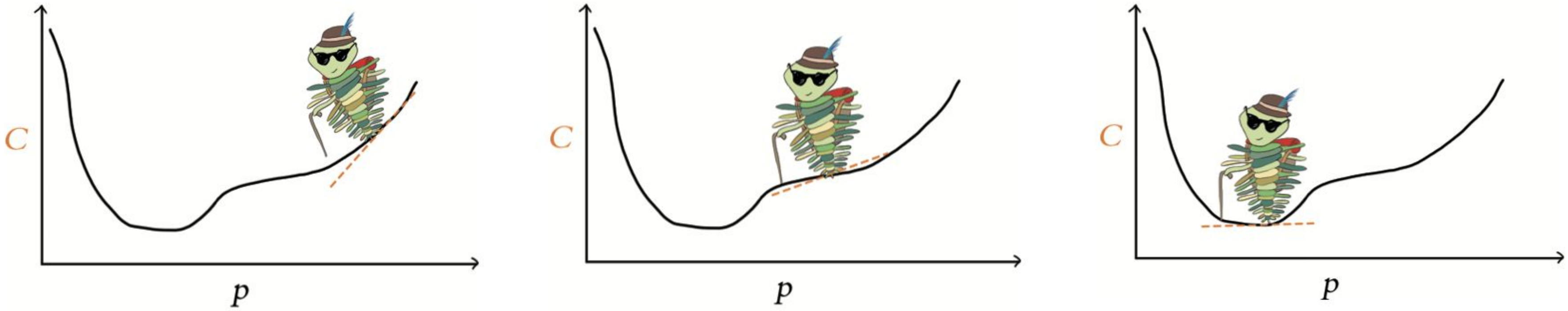
Machine Learning



Machine Learning

Step 3: **Partial Differentiation** (the primary focus of *Calc II*)

Step 4: Descend gradient of cost ***C*** w.r.t. ***p***arameters ***m*** and ***b***



gradient of ***C*** w.r.t. ***p*** = 0

Hands-on code demo: `regression-in-pytorch.ipynb`

Calc II: Partial Derivatives & Integrals

1. Review of Introductory Calculus
2. **Machine Learning Gradients**
3. Integrals

Segment 2: ML Gradients

- Partial Derivatives of Multivariate Functions
- The Partial-Derivative Chain Rule
- Quadratic Cost
- Gradients
- Gradient Descent
- Backpropagation
- Higher-Order Partial Derivatives

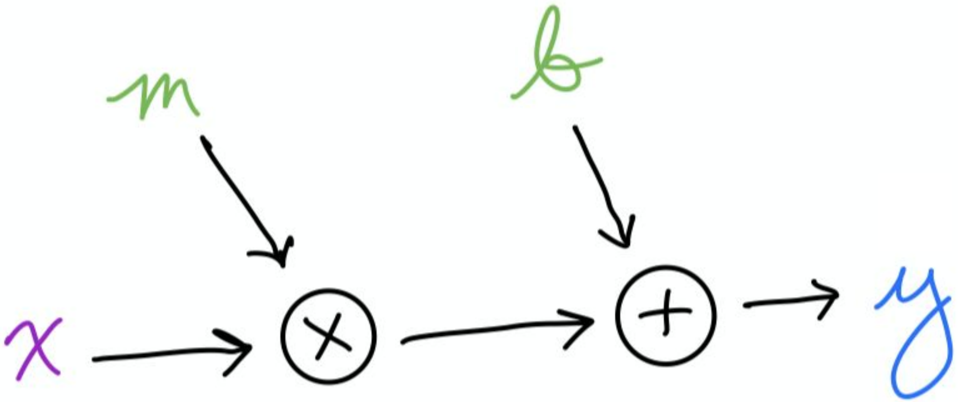
Multivariate Functions

Even in a simple regression such as $y = mx + b$:

y is a function of *multiple* variables

— in this case, m and b.

Therefore, we can't calculate the full derivative dy/dm or dy/db.

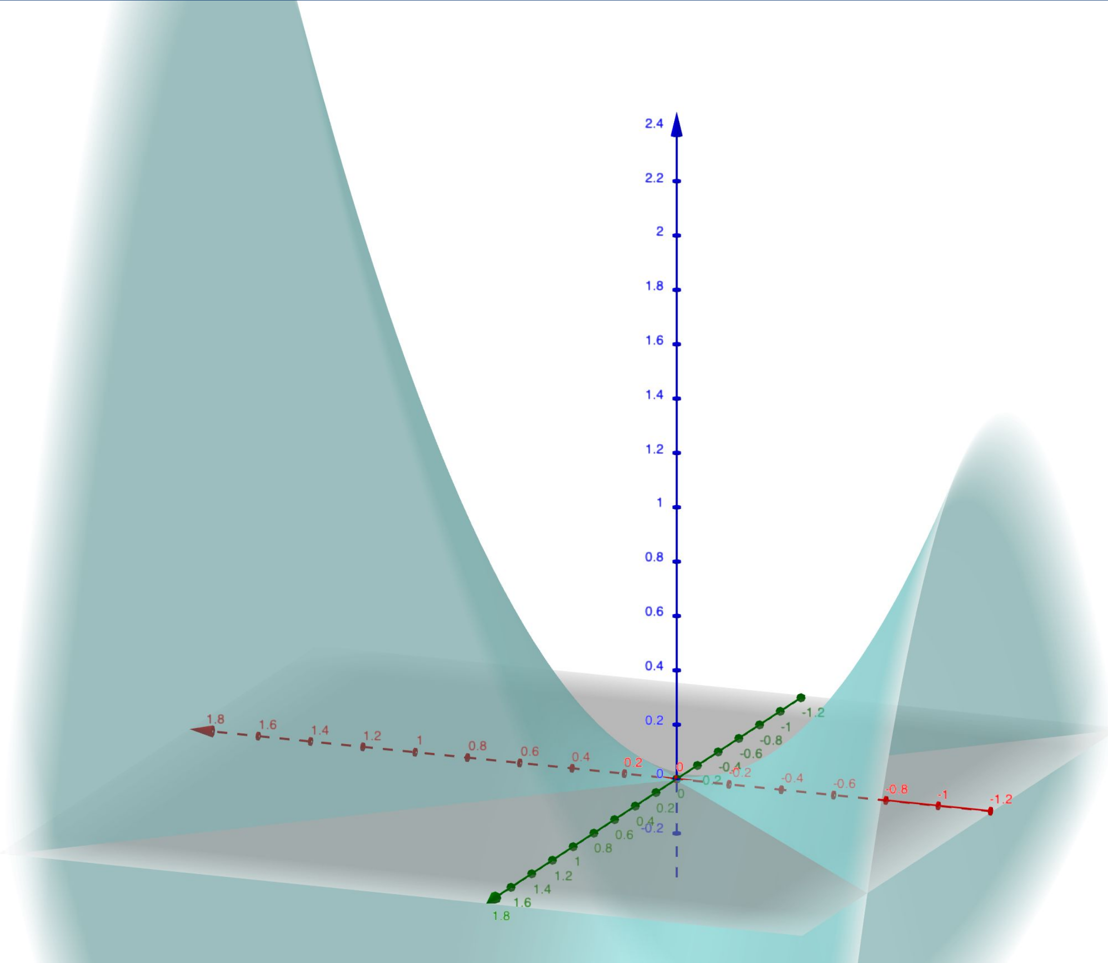


Partial Derivatives

Enable the calculation of derivatives of multivariate equations.

Consider the equation $z = x^2 - y^2$

Hands-on demo:
`geogebra.org/3d`

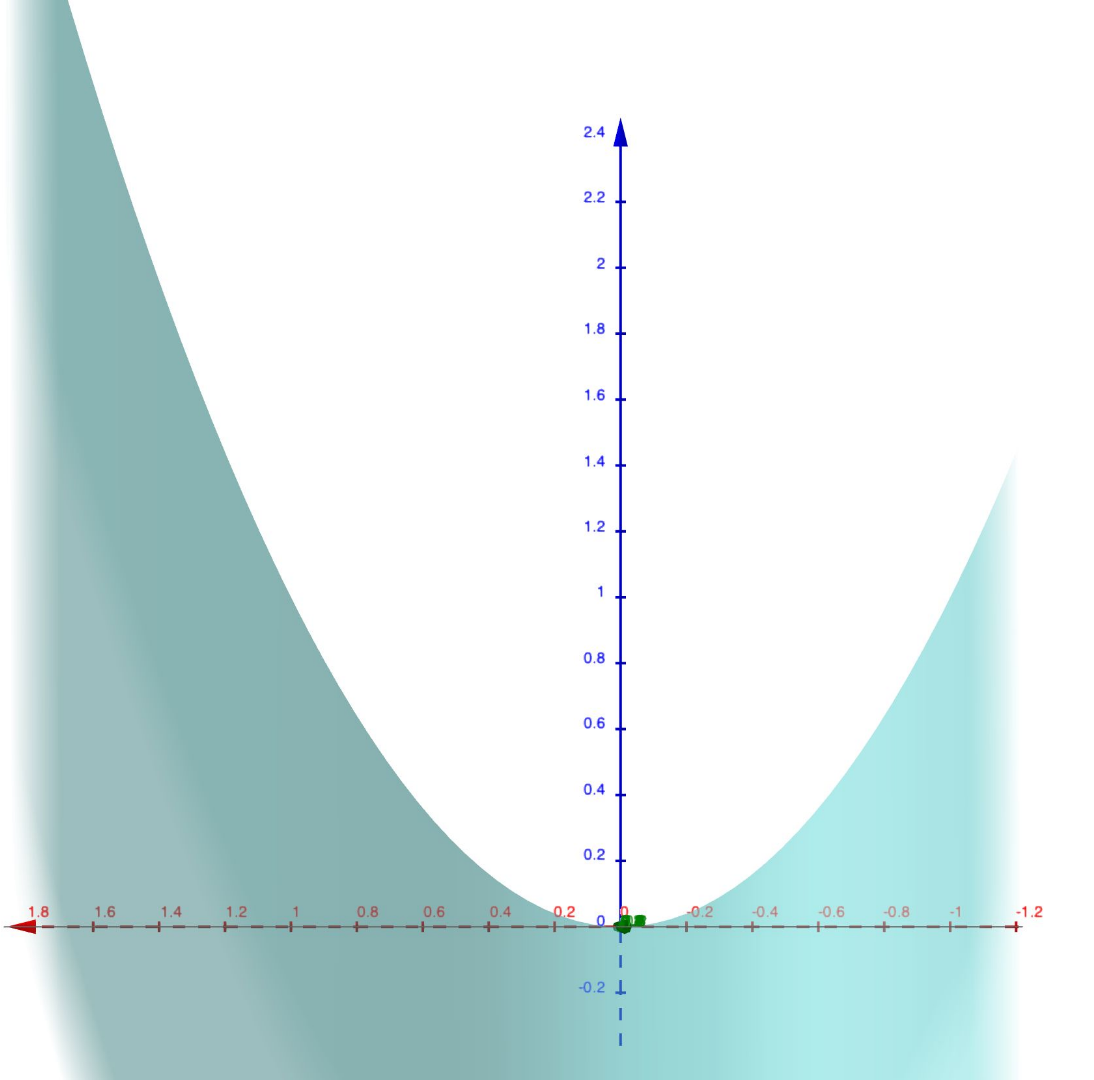


The partial derivative of ***z*** with respect to ***x*** is obtained by considering ***y*** to be a constant:

$$z = x^2 - y^2$$

$$\frac{\partial z}{\partial x} = 2x - 0$$

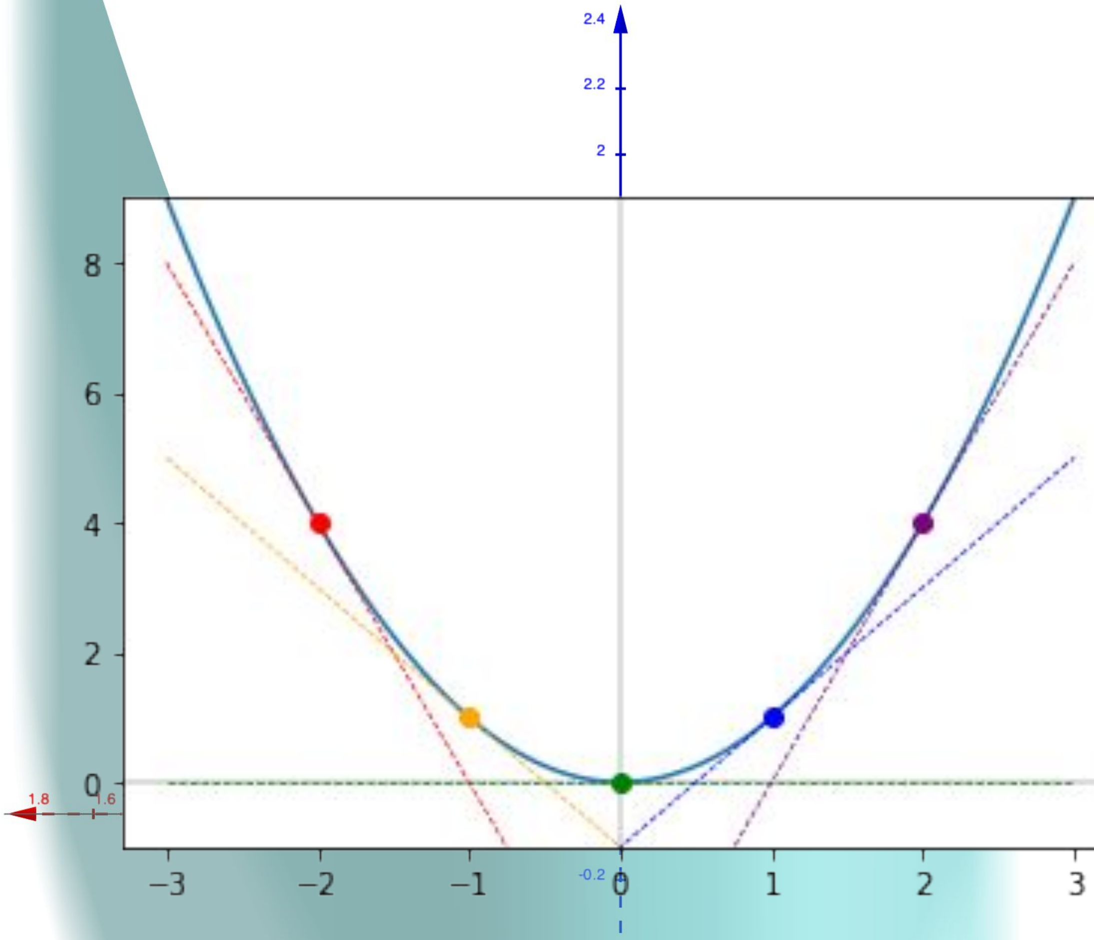
$$= 2x$$



$$\frac{\partial z}{\partial x} = 2x$$

The slope of ***z*** along the ***x*** axis is *twice* the ***x*** axis value.

Hands-on code demo



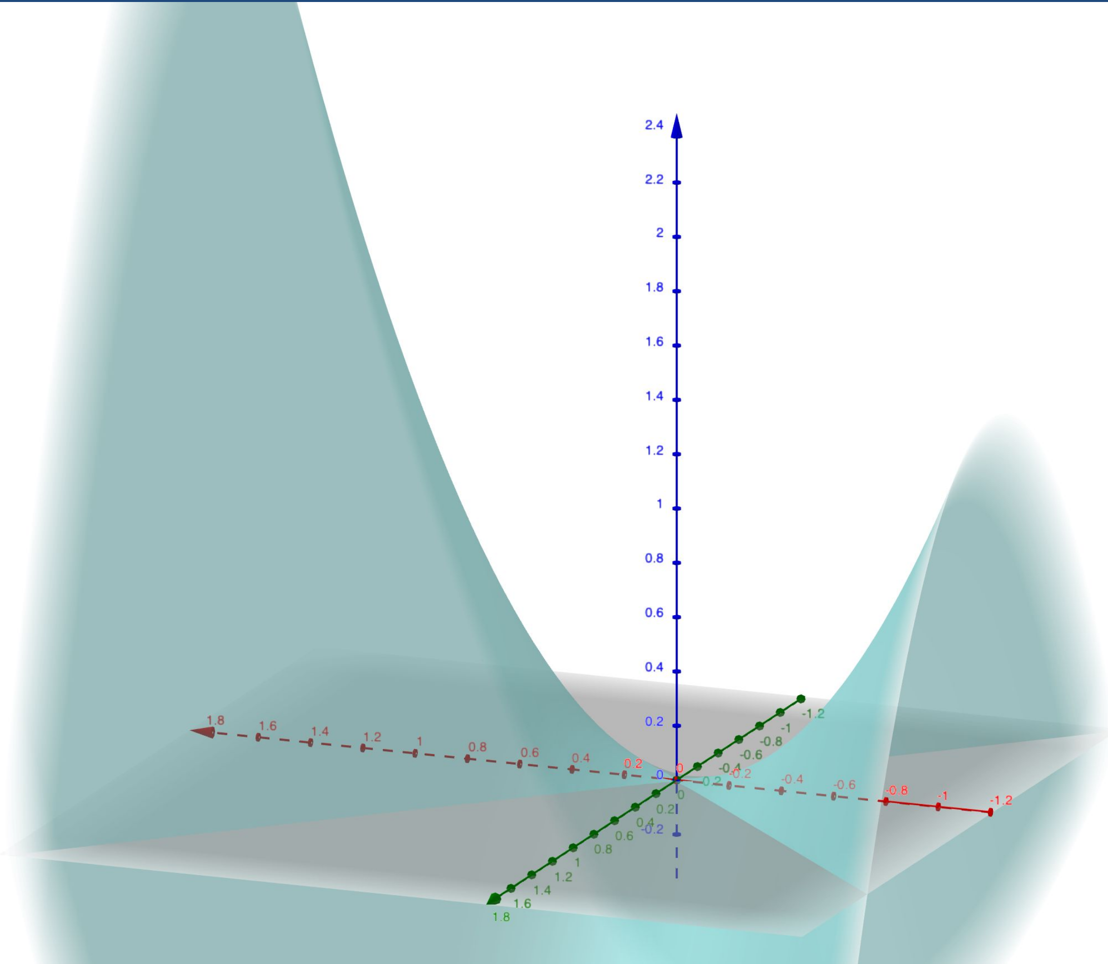
Partial Derivatives

Reconsider $z = x^2 - y^2$

from the perspective of z w.r.t y

Hands-on demo:

geogebra.org/3d

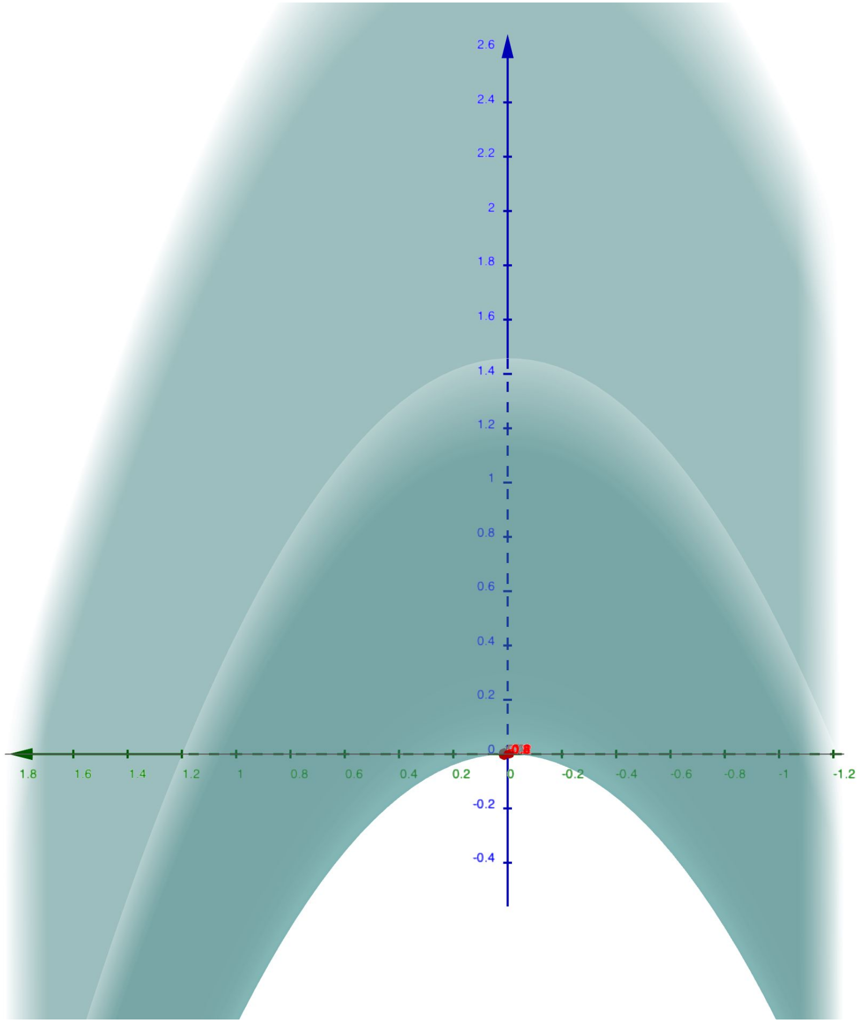


The partial derivative of ***z*** with respect to ***y*** is obtained by considering ***x*** to be a constant:

$$z = x^2 - y^2$$

$$\frac{\partial z}{\partial y} = 0 - 2y$$

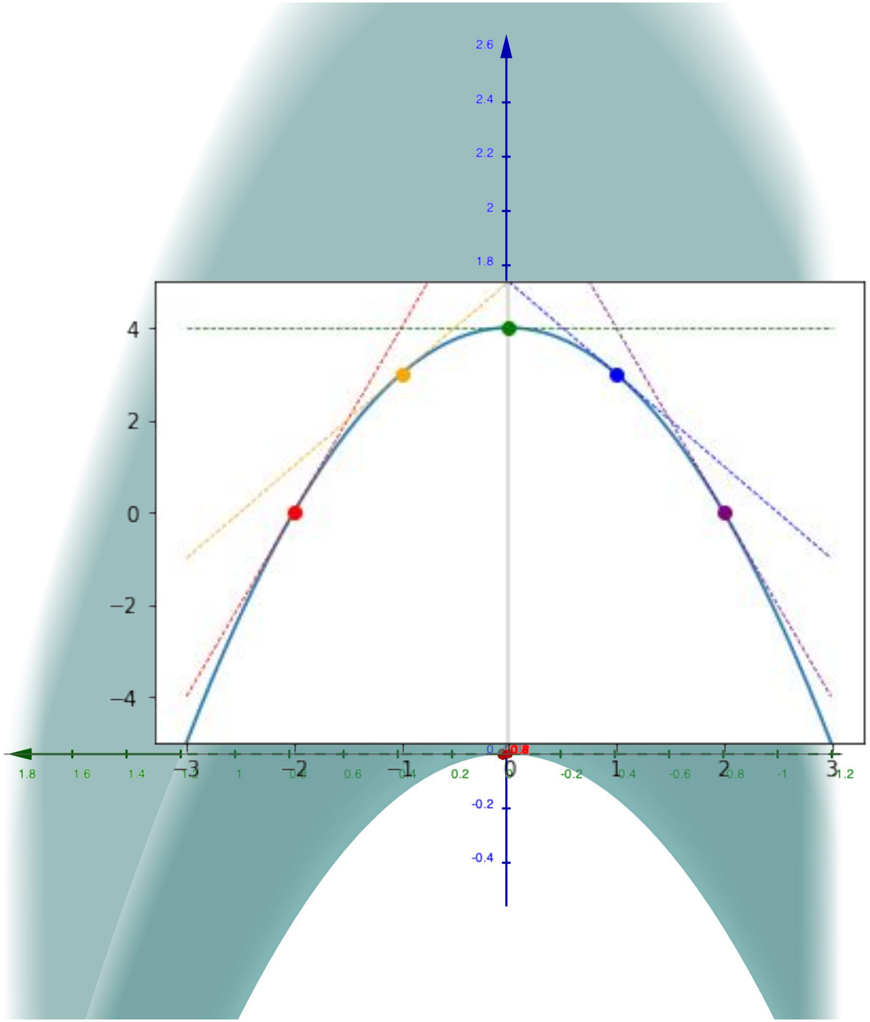
$$= -2y$$



$$\frac{\partial z}{\partial y} = -2y$$

The slope of ***z*** along the ***y*** axis is *twice* the ***y*** axis value ...*and is inverted*.

Hands-on code demo

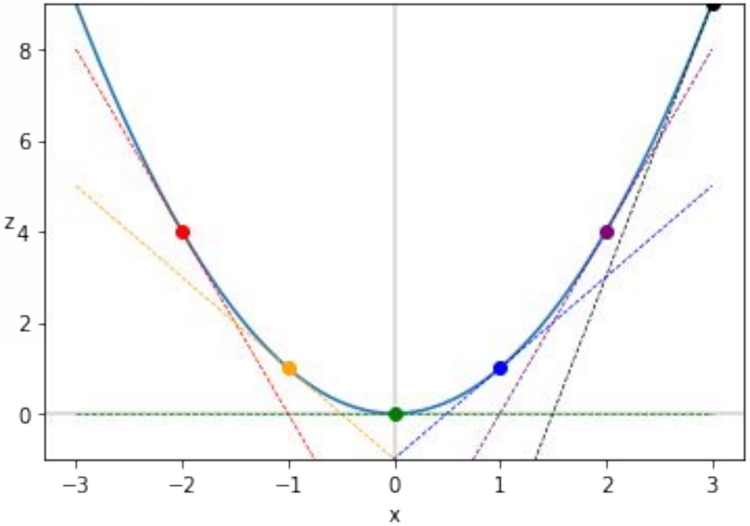


Solutions

① $z = x^2 - y^2 = (3)^2 - (0)^2 = 9 - 0 = 9$

$$\frac{\partial z}{\partial x} = 2x = 2(3) = 6$$

$$\frac{\partial z}{\partial y} = -2y = -2(0) = 0$$

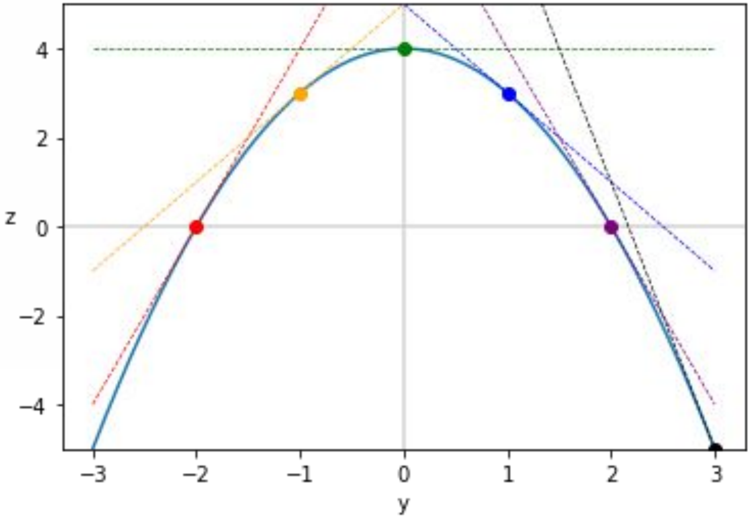


Solutions

② $z = x^2 - y^2 = (2)^2 - (3)^2 = 4 - 9 = -5$

$$\frac{\partial z}{\partial x} = 2x = 2(2) = 4$$

$$\frac{\partial z}{\partial y} = -2y = -2(3) = -6$$



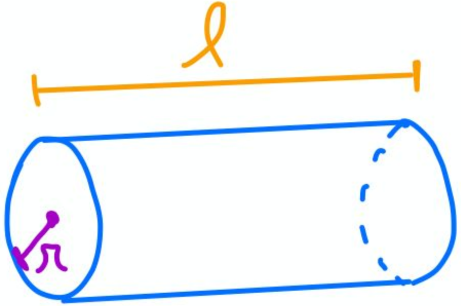
Solutions

(3) $z = x^2 - y^2 = (-2)^2 - (-3)^2 = 4 - 9 = -5$

$$\frac{\partial z}{\partial x} = 2x = 2(-2) = -4$$

$$\frac{\partial z}{\partial y} = -2y = -2(-3) = 6$$

Partial Derivatives



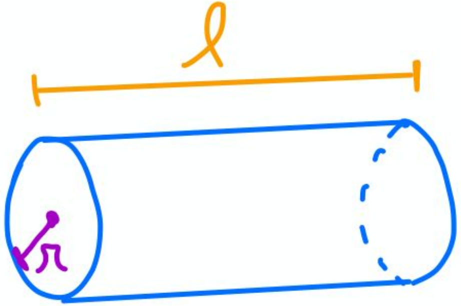
$$v = \pi r^2 \ell$$

$$\frac{\partial v}{\partial \ell} = \pi r^2 \frac{d}{d\ell}(\ell)$$
$$= \pi r^2 (1)$$
$$= \pi r^2$$

$\therefore$ a change in length ℓ corresponds to a change in volume by πr^2

Hands-on code demo

Partial Derivatives



$$v = \pi r^2 \ell$$

$$\frac{\partial v}{\partial r} = \pi \frac{d}{dr}\left(r^2\right)\ell$$

$$= \pi (2r)\ell$$

$$= 2\pi r \ell$$

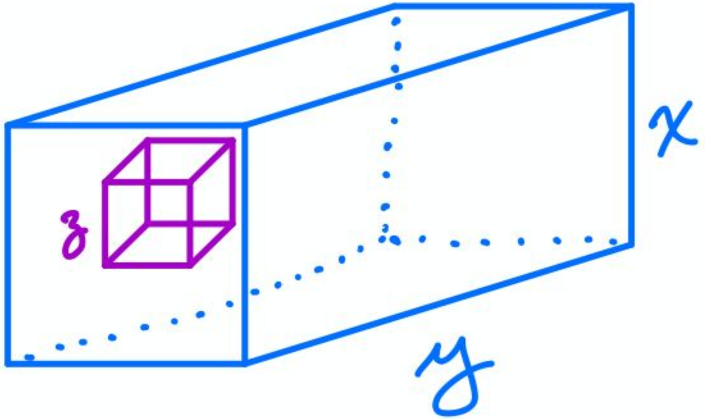
$\therefore$ a change in radius r (by an infinitesimally small amount) corresponds to a change in volume by $2\pi r \ell$

Hands-on code demo

Exercises

Find all the partial derivatives of the following functions:

1. $z = y^3 + 5xy$
2. The surface area of a cylinder is described by $a = 2\pi r^2 + 2\pi rh$.
3. The volume of a **square prism** with a **cube** cut out of its center is described by $v = x^2y - z^3$.



Solutions

1

$$z = y^3 + 5xy$$

$$\frac{\partial z}{\partial y} = 3y^2 + 5x(1)$$

$$= 3y^2 + 5x$$

$$\frac{\partial z}{\partial x} = 0 + 5y(1)$$

$$= 5y$$

Solutions

(2)

$$a = 2\pi r^2 + 2\pi r h$$

$$\frac{\partial a}{\partial r} = 2\pi(2r) + 2\pi h(1)$$
$$= 4\pi r + 2\pi h$$

$$\frac{\partial a}{\partial h} = 0 + 2\pi r(1)$$
$$= 2\pi r$$

Solutions

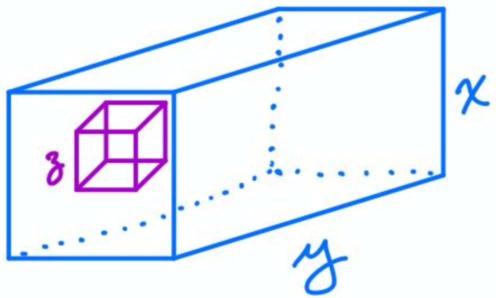
(3)

$$v = x^2 y - z^3$$

$$\frac{\partial v}{\partial x} = 2xy - 0 = 2xy$$

$$\frac{\partial v}{\partial y} = x^2(1) - 0 = x^2$$

$$\frac{\partial v}{\partial z} = 0 - 3z^2 = -3z^2$$



Partial Derivative Notation

$$z = f(x, y)$$

$$\frac{\partial z}{\partial x} \qquad \frac{\partial f}{\partial x} \qquad f_x \qquad D_x f$$

The Chain Rule

Let's say: $y = f(u) \quad u = g(x)$

The Chain Rule

With a multivariate function, the partial derivative is more interesting:

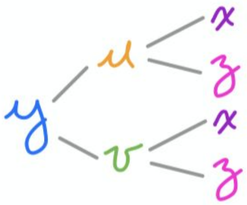
$$y = f(u) \qquad u = g(x, z)$$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} \qquad \frac{\partial y}{\partial z} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial z}$$

The Chain Rule

With multiple multivariate functions, it gets really interesting:

$$y = f(u, v) \qquad u = g(x, z) \qquad v = h(x, z)$$



$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial v}{\partial x}$$

$$\frac{\partial y}{\partial z} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial z} + \frac{\partial y}{\partial v}\frac{\partial v}{\partial z}$$

The Chain Rule

Generalizing completely:

$$y = f(u_1, u_2, \ldots, u_m) \qquad u_j = g(x_1, x_2, \ldots, x_n)$$

for $i = 1, 2, \ldots, n$:

$$\frac{\partial y}{\partial x_i} = \frac{\partial y}{\partial u_1}\frac{\partial u_1}{\partial x_i} + \frac{\partial y}{\partial u_2}\frac{\partial u_2}{\partial x_i} + \ldots + \frac{\partial y}{\partial u_m}\frac{\partial u_m}{\partial x_i}$$

Exercises

Find all the partial derivatives of *y*, where:

1. $y = f(u, v) \quad u = g(x) \quad v = h(z)$

2. $y = f(u, v) \quad u = g(x) \quad v = h(x, z)$

3. $y = f(u, v, w) \quad u = g(x) \quad v = h(x) \quad w = j(x)$

Solutions

① $y = f(u, v) \quad u = g(x) \quad v = h(z)$

$$\begin{tikzcd}[row sep=small] & u \arrow[r, dash] & x \\ y \arrow[ur, dash] \arrow[dr, dash] & & \\ & v \arrow[r, dash] & z \end{tikzcd}$$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} \qquad \frac{\partial y}{\partial z} = \frac{\partial y}{\partial v}\frac{\partial v}{\partial z}$$

Solutions

② $y = f(u, v) \quad u = g(x) \quad v = h(x, z)$

$$\begin{tikzcd}[arrows=dash]
 & u \arrow[r] & x \\
y \arrow[ur] \arrow[dr] & & x \\
 & v \arrow[ur] \arrow[dr] & \\
 & & z
\end{tikzcd}$$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial v}{\partial x} \qquad \frac{\partial y}{\partial z} = \frac{\partial y}{\partial v}\frac{\partial v}{\partial z}$$

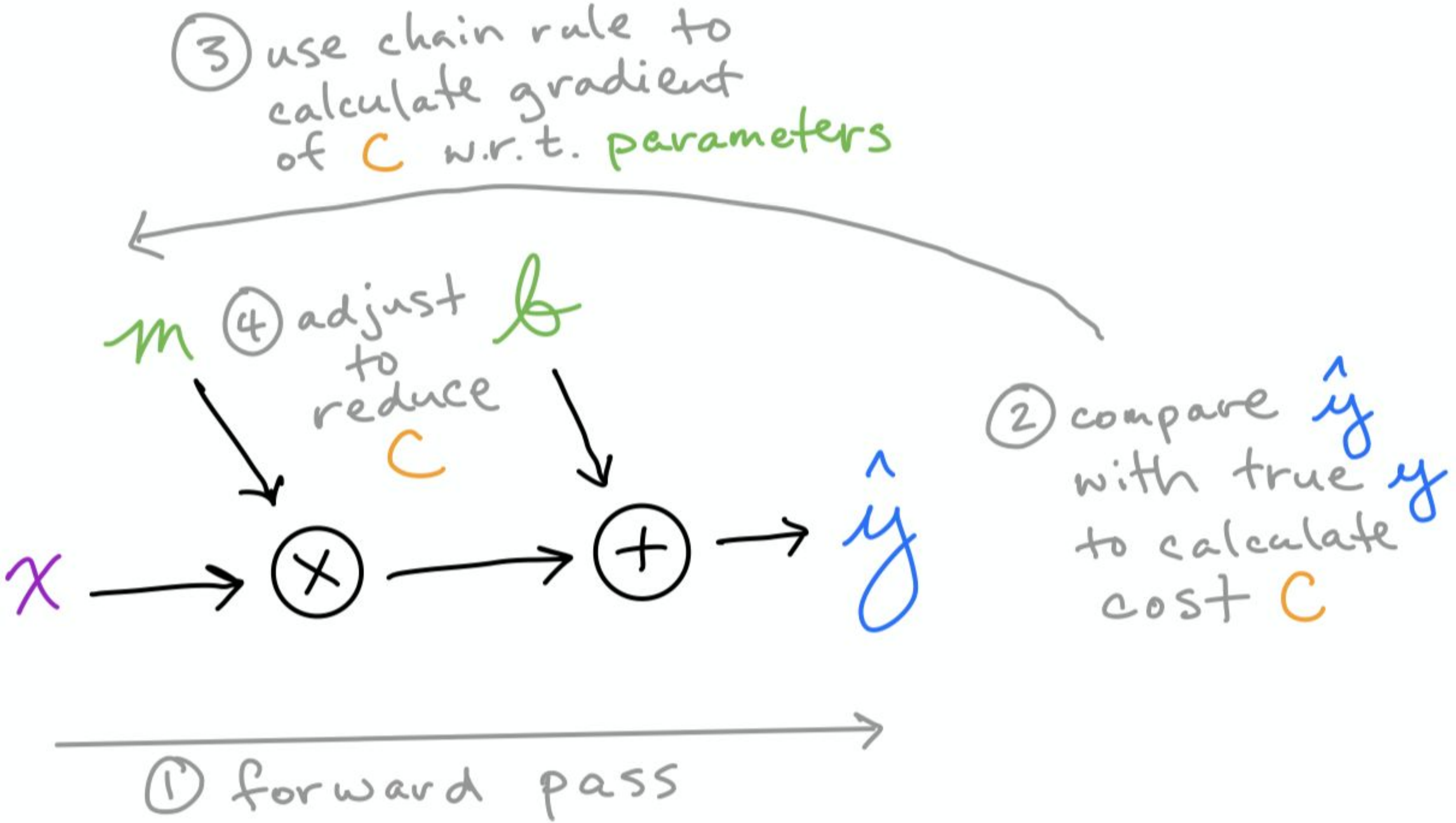
Solutions

(3) $y = f(u, v, w) \quad u = g(x) \quad v = h(x) \quad w = j(x)$

$$
\begin{tikzcd}
 & u \arrow[r, dash] & x \\
y \arrow[ur, dash] \arrow[r, dash] \arrow[dr, dash] & v \arrow[r, dash] & x \\
 & w \arrow[r, dash] & x
\end{tikzcd}
$$

$$\frac{dy}{dx} = \frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial y}{\partial w}\frac{\partial w}{\partial x}$$

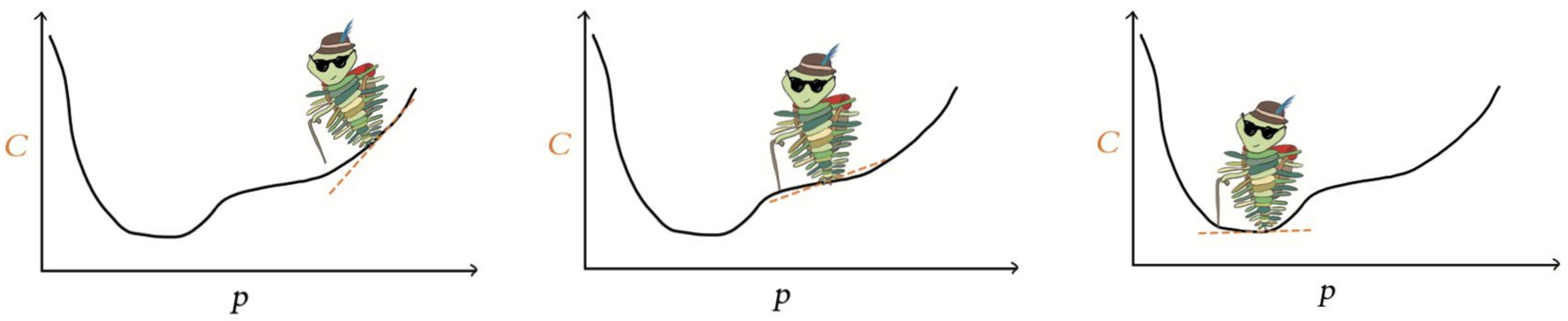
Recalling Machine Learning



Recalling Machine Learning

Step 3: Automatic differentiation

Step 4: Descend gradient of cost ***C*** w.r.t. ***p***arameters ***m*** and ***b***



gradient of ***C*** w.r.t. ***p*** = 0

Hands-on code demo: `single-point-regression-gradient NB`

Quadratic Cost w.r.t. Predicted y

$$C = (\hat{y} - y)^2$$

$$C = u^2 \qquad u = \hat{y} - y$$

$$\frac{dC}{du} = \frac{\partial C}{\partial u} = 2u = 2(\hat{y} - y) \qquad \frac{\partial u}{\partial \hat{y}} = 1 - 0 = 1$$

$$\frac{\partial C}{\partial \hat{y}} = \frac{\partial C}{\partial u}\frac{\partial u}{\partial \hat{y}} = (2(\hat{y} - y))(1) = 2(\hat{y} - y)$$

Predicted y w.r.t. Model Parameters

$$\hat{y} = mx + b$$

$$\frac{\partial \hat{y}}{\partial m} = (1)x + 0$$

$$= x$$

$$\frac{\partial \hat{y}}{\partial b} = 0 + 1$$

$$= 1$$

Quadratic Cost w.r.t. Model Parameters

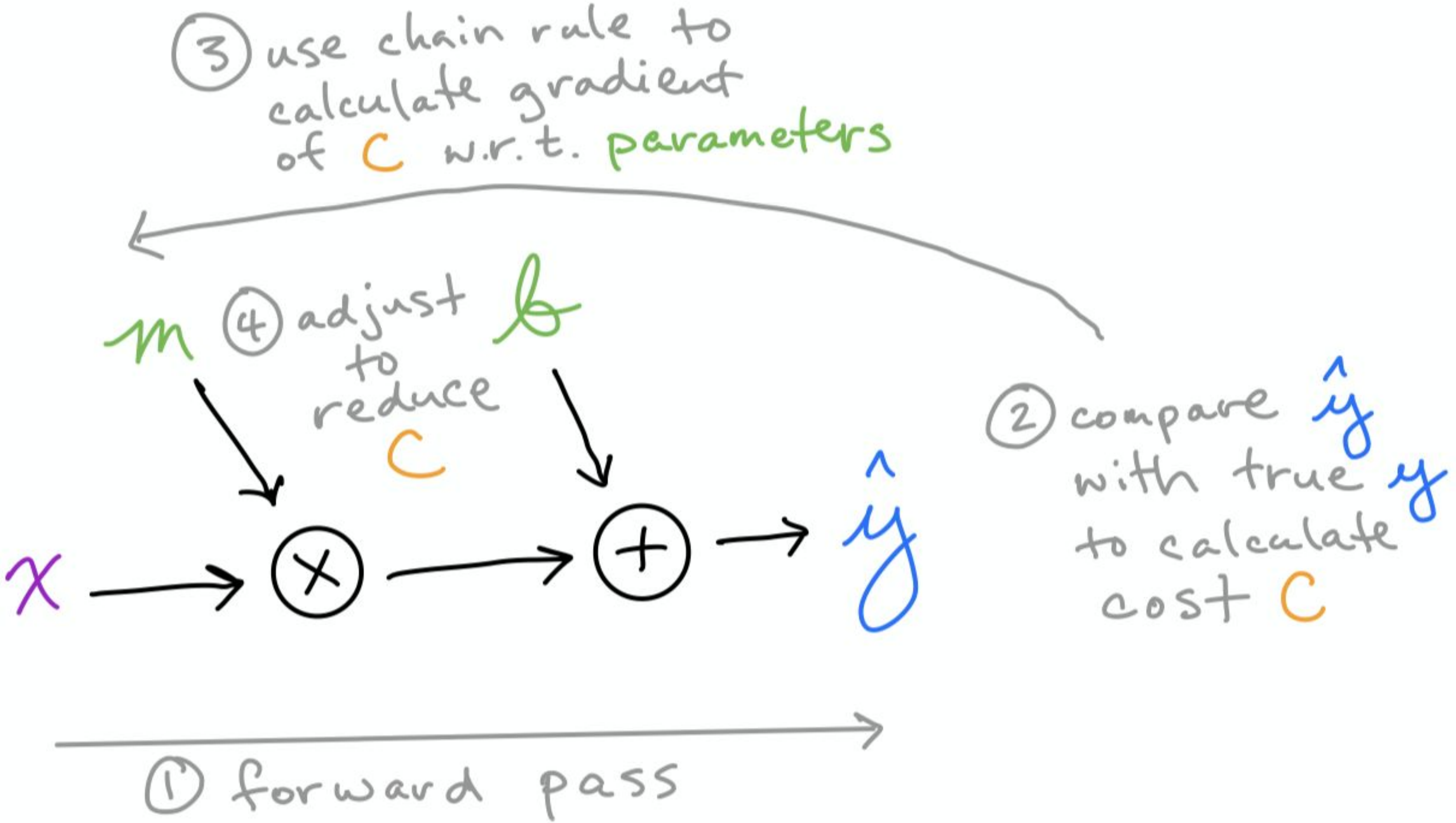
$$\frac{\partial C}{\partial \hat{y}} = 2(\hat{y} - y) \qquad \frac{\partial \hat{y}}{\partial m} = x \qquad \frac{\partial \hat{y}}{\partial b} = 1$$

$$\frac{\partial C}{\partial m} = \frac{\partial C}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial m} = \left(2(\hat{y} - y)\right)x = 2x(\hat{y} - y)$$

$$\frac{\partial C}{\partial b} = \frac{\partial C}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b} = \left(2(\hat{y} - y)\right)1 = 2(\hat{y} - y)$$

Hands-on code demo

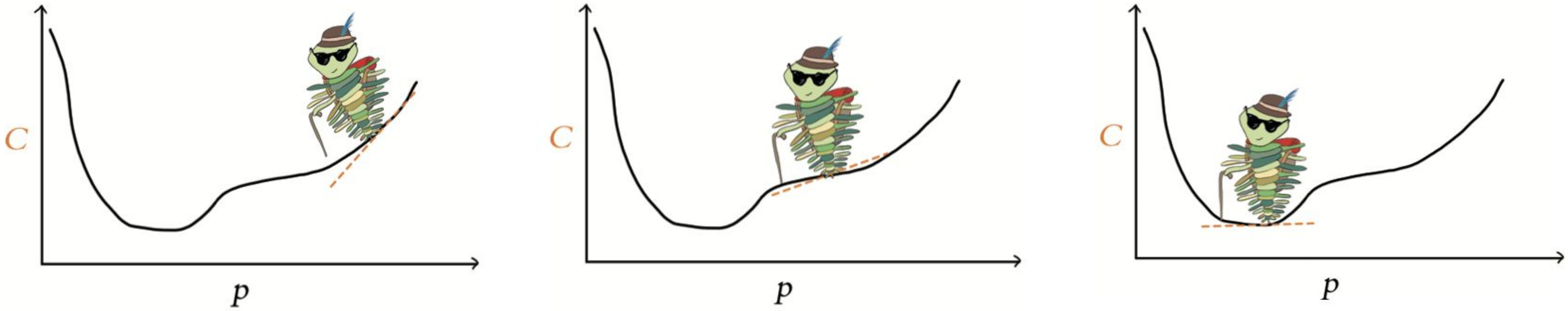
Recalling Machine Learning



Recalling Machine Learning

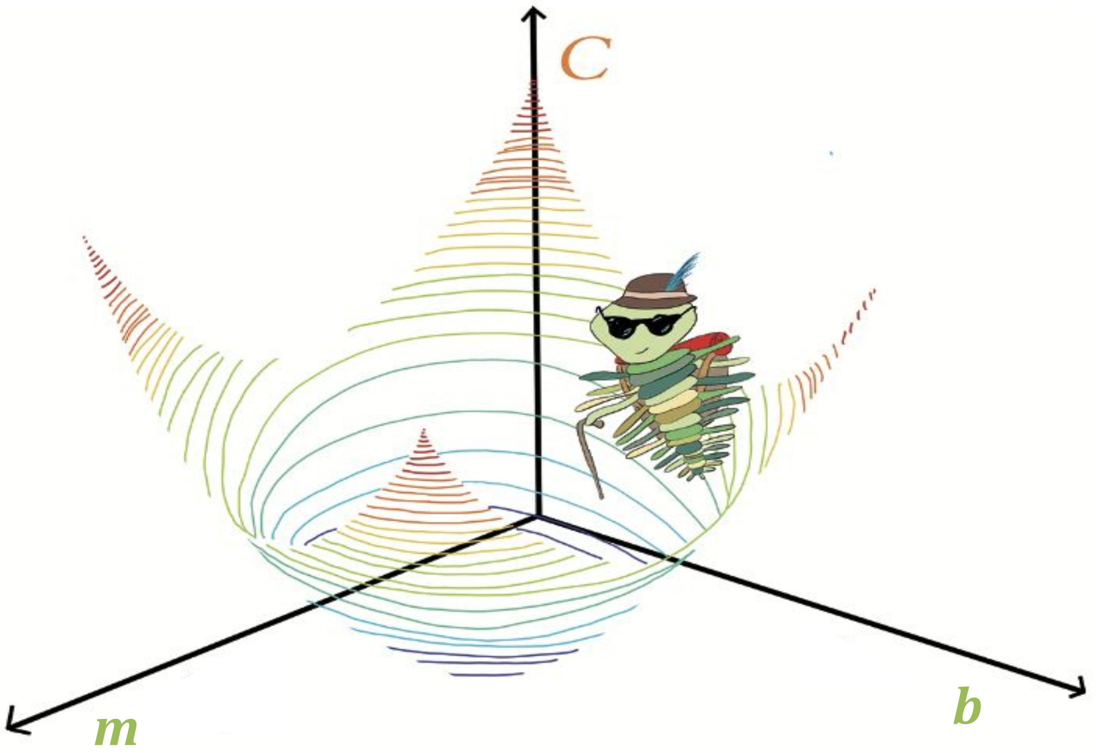
Step 3: Determine gradient of cost ***C*** w.r.t. ***p***arameters ***m*** and ***b***

Step 4: <u>Descend gradient</u>



gradient of ***C*** w.r.t. ***p*** = 0

∇C: the Gradient of Cost



Hands-on code demo: `batch-regression-gradient` NB

MSE w.r.t. Predicted y

$$C = \frac{1}{n}\sum(\hat{y}_i - y_i)^2$$

$$C = \frac{1}{n}\sum u^2$$

$$u = \hat{y}_i - y_i$$

$$\frac{\partial C}{\partial u} = \frac{1}{n}\sum 2u$$

$$\frac{\partial u}{\partial \hat{y}_i} = 1 - 0 = 1$$

$$= \frac{2}{n}\sum u$$

$$= \frac{2}{n}\sum(\hat{y}_i - y_i)$$

MSE w.r.t. Model Parameters

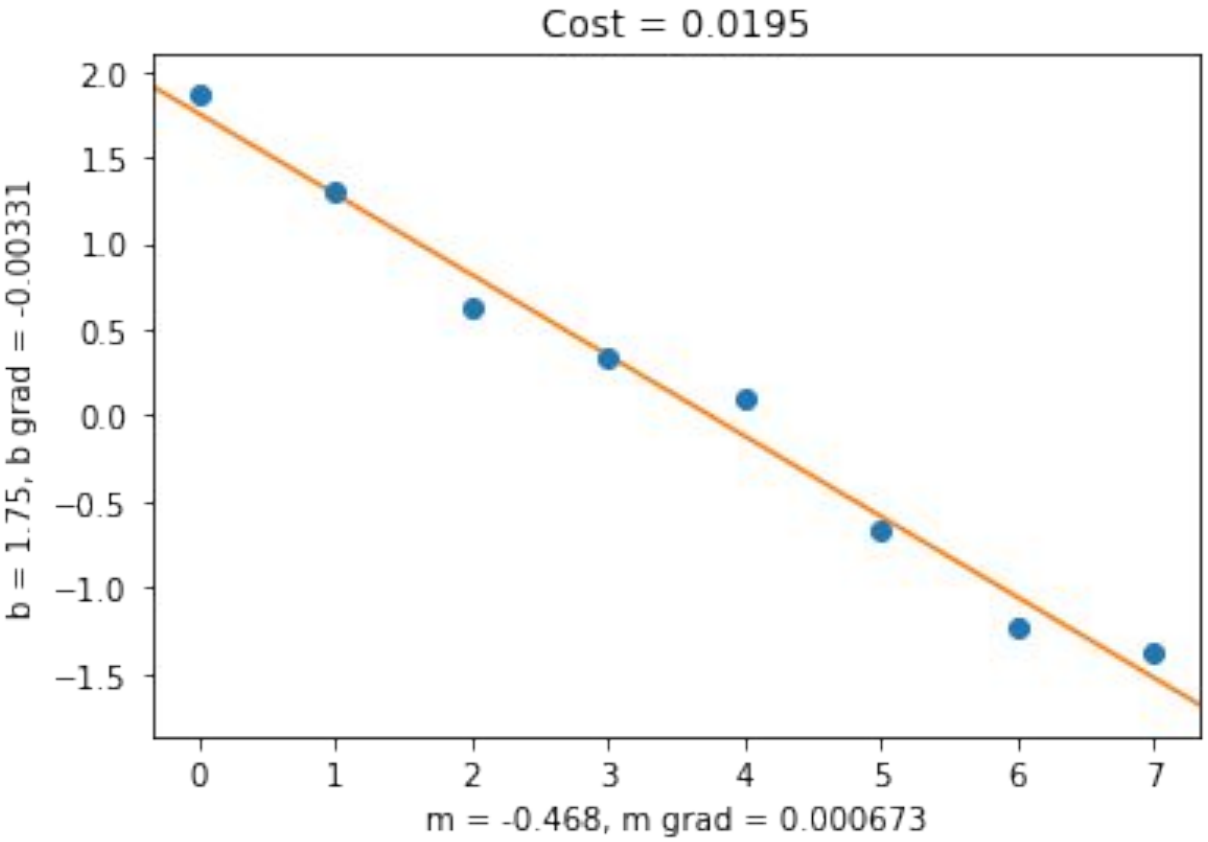
$$\frac{\partial C}{\partial u} = \frac{2}{n}\sum(\hat{y}_i - y_i) \qquad \frac{\partial u}{\partial \hat{y}_i} = 1 \qquad \frac{\partial \hat{y}_i}{\partial m} = x_i \qquad \frac{\partial \hat{y}_i}{\partial b} = 1$$

$$\frac{\partial C}{\partial b} = \frac{\partial C}{\partial u}\frac{\partial u}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial b} = \frac{2}{n}\sum(\hat{y}_i - y_i)\cdot 1 \cdot 1 = \frac{2}{n}\sum(\hat{y}_i - y_i)$$

$$\frac{\partial C}{\partial m} = \frac{\partial C}{\partial u}\frac{\partial u}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial m} = \frac{2}{n}\sum(\hat{y}_i - y_i)\cdot 1 \cdot x_i = \frac{2}{n}\sum(\hat{y}_i - y_i)\cdot x_i$$

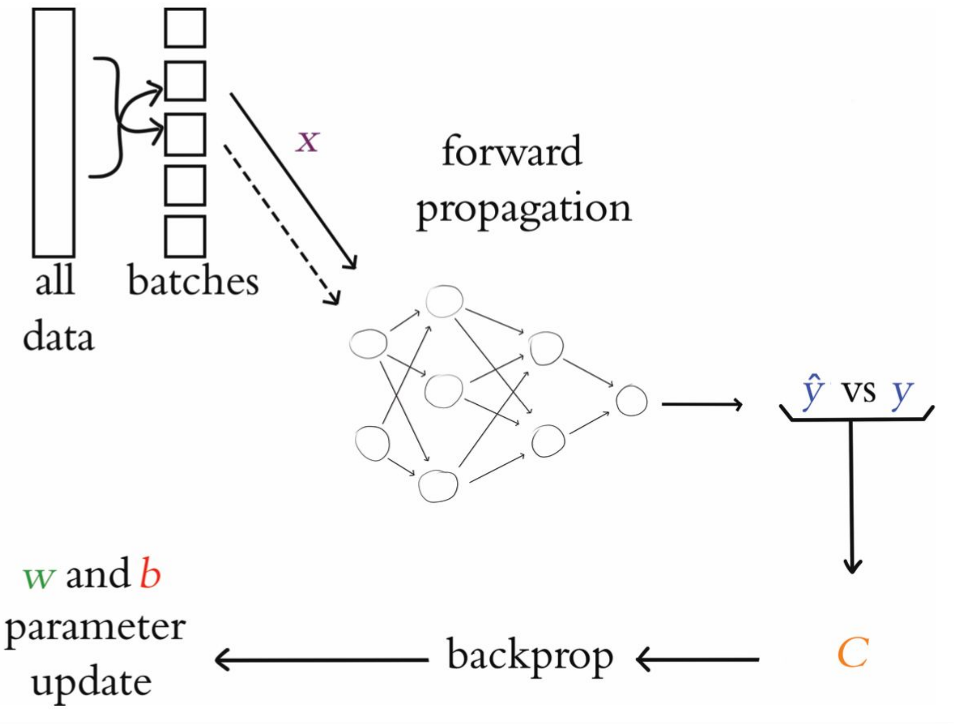
Hands-on code demo

Regression Line after 1000 Epochs

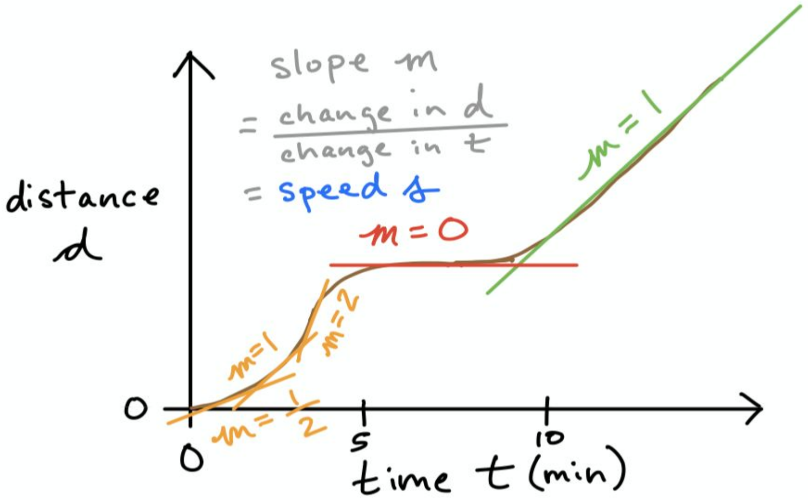


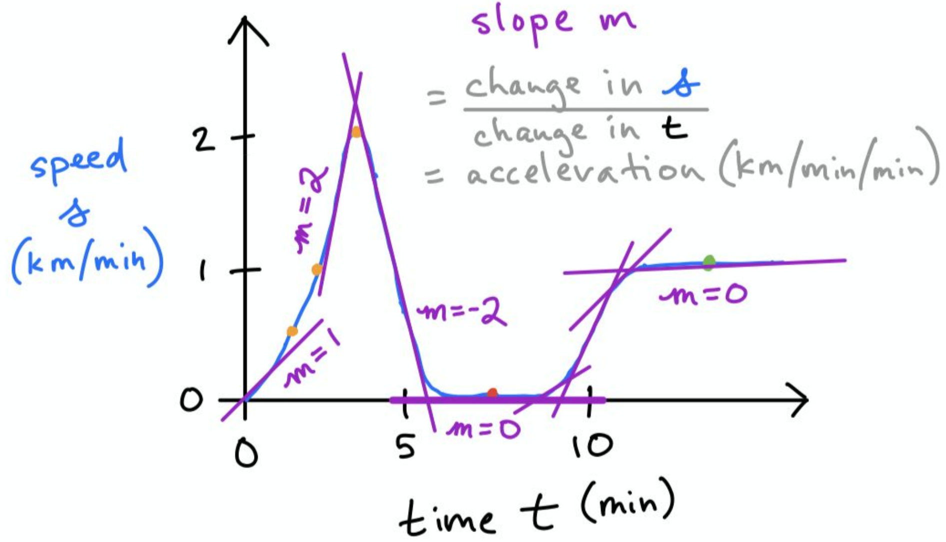
Backpropagation

Chain rule of partial derivatives of **cost** w.r.t. model parameters extends to deep neural networks, which may have 1000s of layers:



Higher-Order Derivatives





Higher-Order Partial Derivatives

In ML, used to accelerate through gradient descent. (*Optimization*)

Consider the following first-order partial derivatives...

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

$$\frac{\partial z}{\partial x} = 2x + 5y(1) + 0 \quad = 2x + 5y$$

$$\frac{\partial z}{\partial y} = 0 + 5x(1) + 2(2y) = 5x + 4y$$

Higher-Order Partial Derivatives

$$\frac{\partial z}{\partial x} = 2x + 5y \qquad \frac{\partial z}{\partial y} = 5x + 4y$$

Unmixed:

$$\frac{\partial^2 z}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial x}(2x + 5y) = 2(1) + 0 = 2$$

$$\frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial}{\partial y}(5x + 4y) = 0 + 4(1) = 4$$

Higher-Order Partial Derivatives

$$\frac{\partial z}{\partial x} = 2x + 5y \qquad \frac{\partial z}{\partial y} = 5x + 4y$$

Mixed:

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial}{\partial x}(5x + 4y) = 5(1) + 0 = 5$$

$$\frac{\partial^2 z}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial y}(2x + 5y) = 0 + 5(1) = 5$$

Clairaut's (or Schwartz's, or Young's) Theorem:

Under certain (common) conditions, $\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}$

Higher-Order Partial Derivative Notation

$$z = f(x, y)$$

$$\frac{\partial^2 z}{\partial x^2} \qquad \frac{\partial^2 f}{\partial x^2} \qquad f_{xx} \qquad D_{xx} f$$

$$\frac{\partial^2 z}{\partial y \partial x} \qquad \frac{\partial^2 f}{\partial y \partial x} \qquad f_{xy} \qquad D_{xy} f$$

Exercise

Find all the second-order partial derivatives of $z = x^3 + 2xy$.

Solution

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

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

$$\frac{\partial z}{\partial y} = 2x$$

$$\frac{\partial^2 z}{\partial x^2} = 6x$$

$$\frac{\partial^2 z}{\partial y^2} = 0$$

$$\frac{\partial^2 z}{\partial y \partial x} = 2$$

$$\frac{\partial^2 z}{\partial x \partial y} = 2$$

Calc II: Partial Derivatives & Integrals

1. Review of Introductory Calculus
2. Gradients Applied to Machine Learning
3. **Integrals**

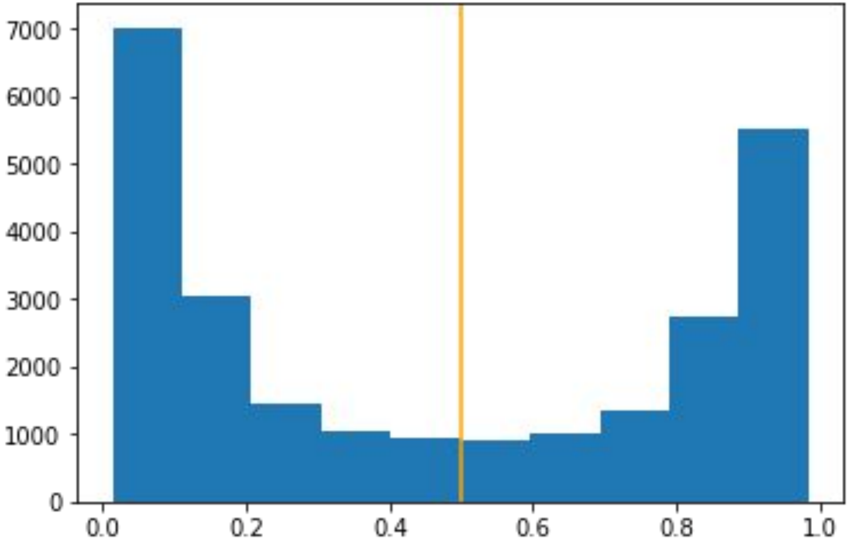
Segment 3: Integrals

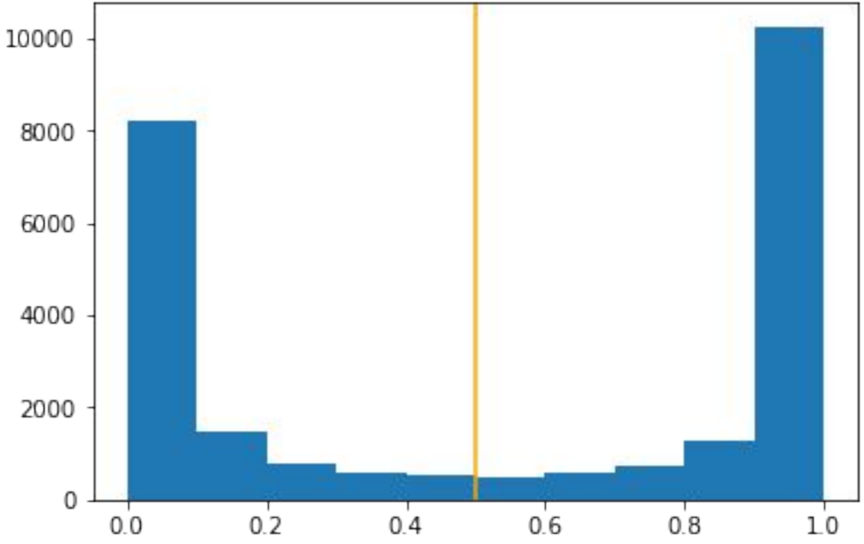
- Binary Classification
- The Confusion Matrix
- The Receiver-Operating Characteristic (ROC) Curve
- Calculating Integrals Manually
- Numeric Integration with Python
- Finding the Area Under the ROC Curve
- Resources for Further Study of Calculus

Supervised Learning

- Have x and y
- Goal: learn function that uses x to approximate y
- Examples:
    - **Regression**
        - Clinical measure of forgetfulness
        - Sales of a product
        - Future value of an asset
    - **Classification**
        - ***Multinomial***
            - Handwritten digits: 10 classes
            - Imagenet: 21k classes
        - ***Binomial***
            - Movie-review sentiment: positive vs negative
            - Photos of fast food: Hot dog vs not hot dog

Accuracy at a Single Threshold



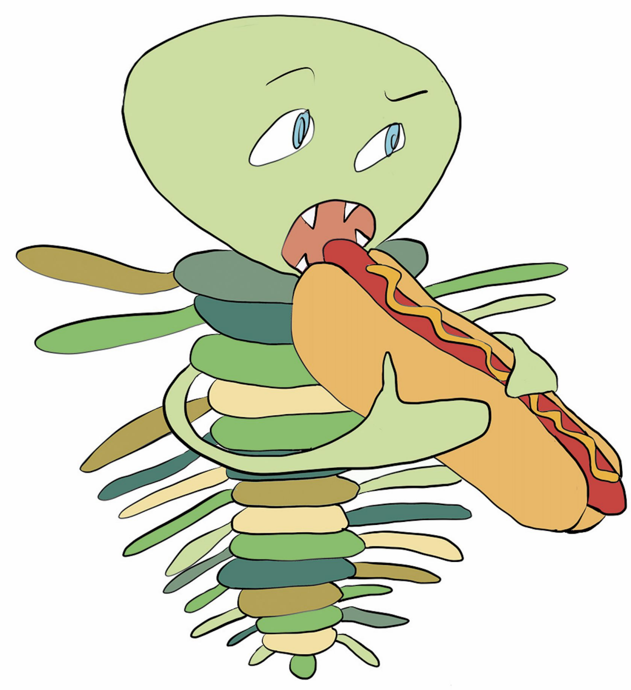


- Doesn’t reflect model quality at other points in output distribution
- If y = 1: Prediction of 0.49 is 100% wrong; 0.51 prediction 100% correct
  - Prediction of 0.51 is considered as correct as prediction of 0.99
- Solution: ROC AUC metric

The Confusion Matrix

		actual *y*	
		1	0
predicted *y*	1	True positive	False positive
	0	False negative	True negative

Four Hot Dog Predictions



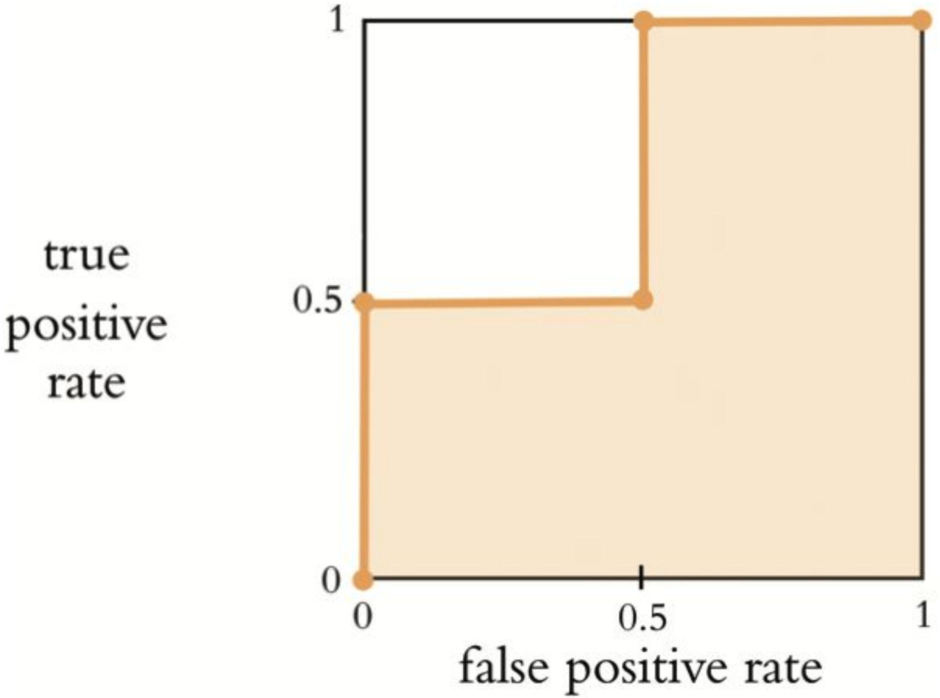
y	$\hat{y}$

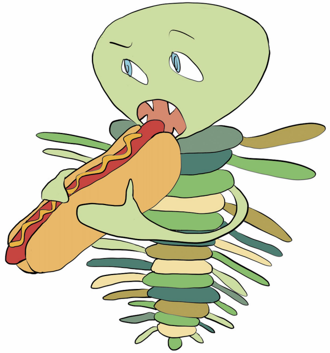
JonKrohn.com

Receiver-Operating Characteristic

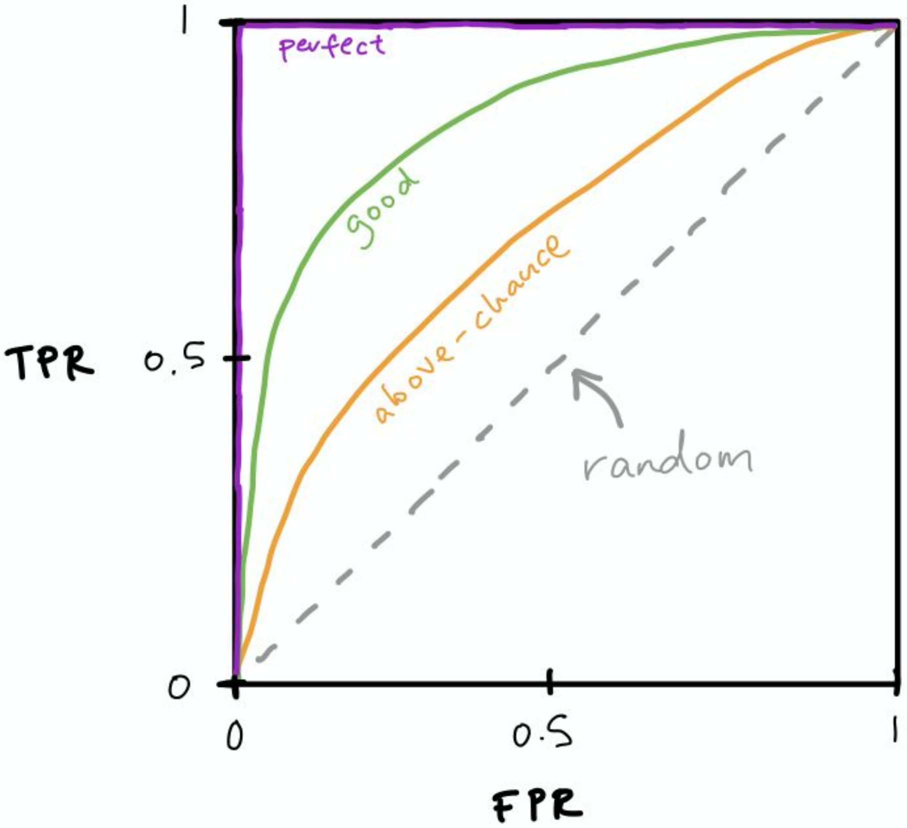
y	$\hat{y}$	0.3 threshold	0.5 threshold	0.6 threshold
0 (not hot dog)	0.3			
1 (hot dog)	0.5			
0 (not hot dog)	0.6			
1 (hot dog)	0.9			

The ROC Curve



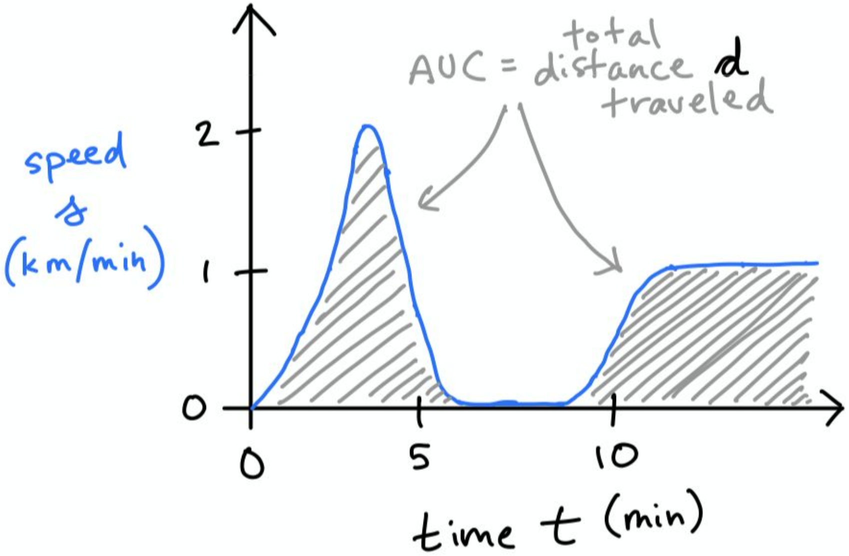


The ROC Curve



Integral Calculus

- **Study of areas under curves**
- Facilitates the inverse of differential calculus:

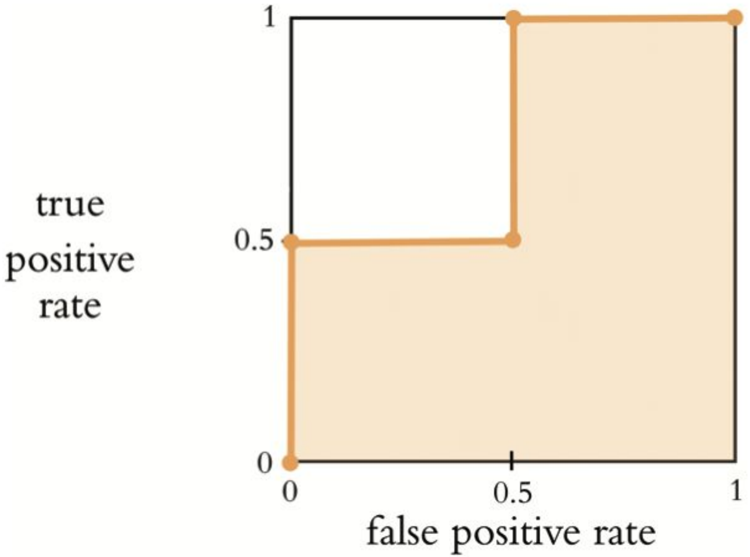


- Also finds areas more generally, volumes, and central points

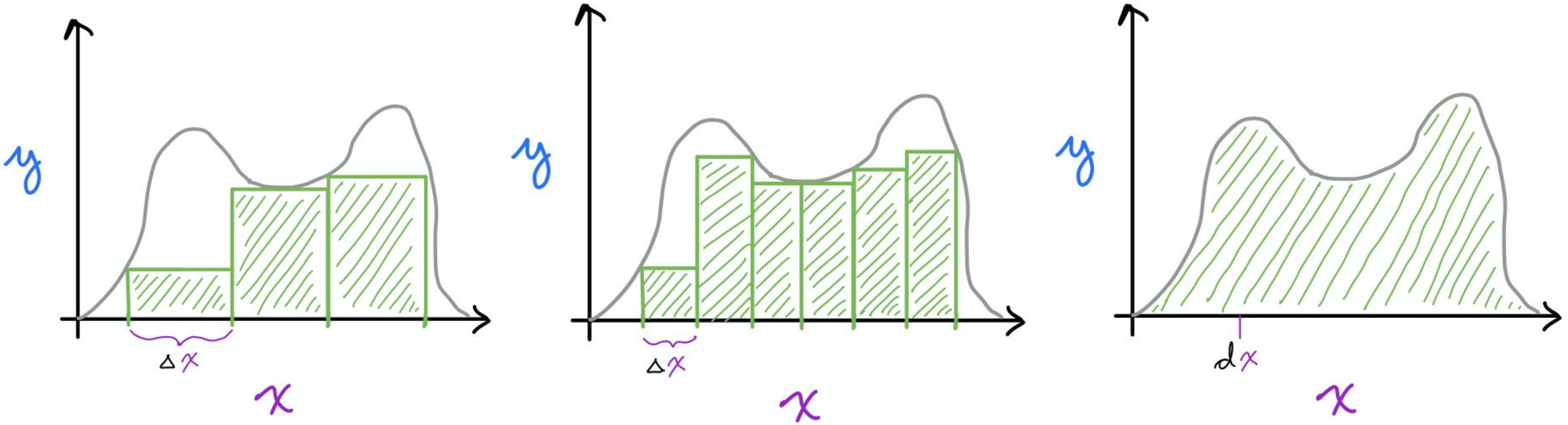
Integral Calculus Applications in ML

Find area under the curve:

- **Receiver operating characteristic** *(Calc II)*
- Probability theory's "expectation" of random variable is widely used in machine learning, incl. deep learning *(Prob. & Info. Thy.)*



dx Slice Width



dx indicates slice width (Δ*x*) is approaching zero width

Integral Notation

integral symbol (indefinite)

$$\int \underbrace{2x}_{\text{function to integrate}} \; \underbrace{dx}_{\text{variable to integrate along}}$$

The Power Rule

$$\int x^n \, dx = \frac{x^{n+1}}{n+1} + C$$

constant that may have been lost by differentiation

note $n \neq -1$

The Constant Multiple Rule

$$\int c f(x)\,dx = c\int f(x)\,dx + C$$

The Sum Rule

$$\int (f(x) + g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx + C$$

Exercises

1. $\int x^5 \, dx$

2. $\int 12x^5 \, dx$

3. $\int (12x^5 - x) \, dx$

Solutions

① $\int x^5 \, dx = \frac{x^6}{6} + C$

② $\int 12x^5 \, dx = \frac{12x^6}{6} + C = 2x^6 + C$

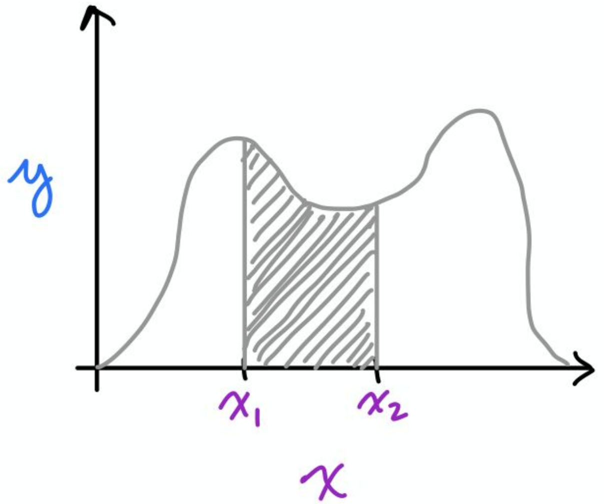
③ $\int (12x^5 - x) \, dx = 2x^6 - \frac{x^2}{2} + C$

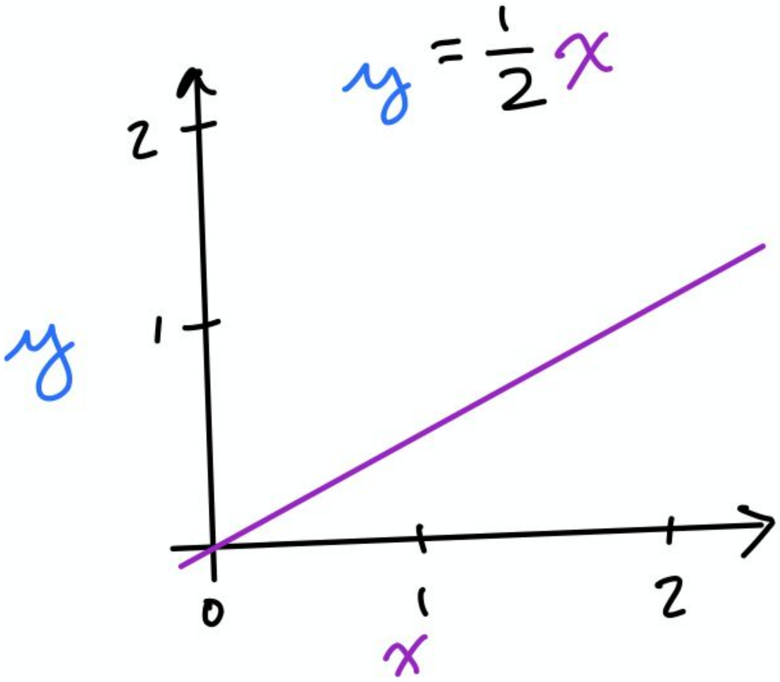
Definite Integrals

$$\int_{x_1}^{x_2} f(x)\,dx$$

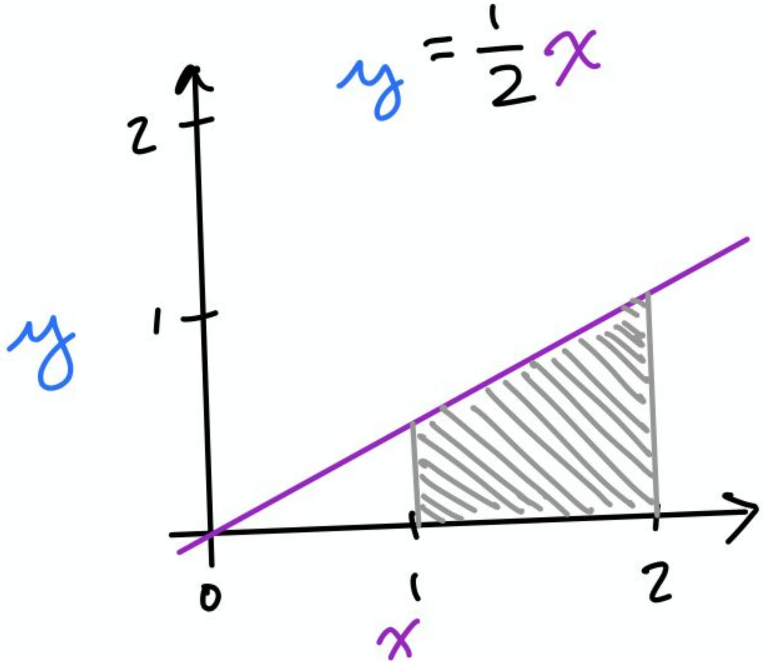
start of range

end of range

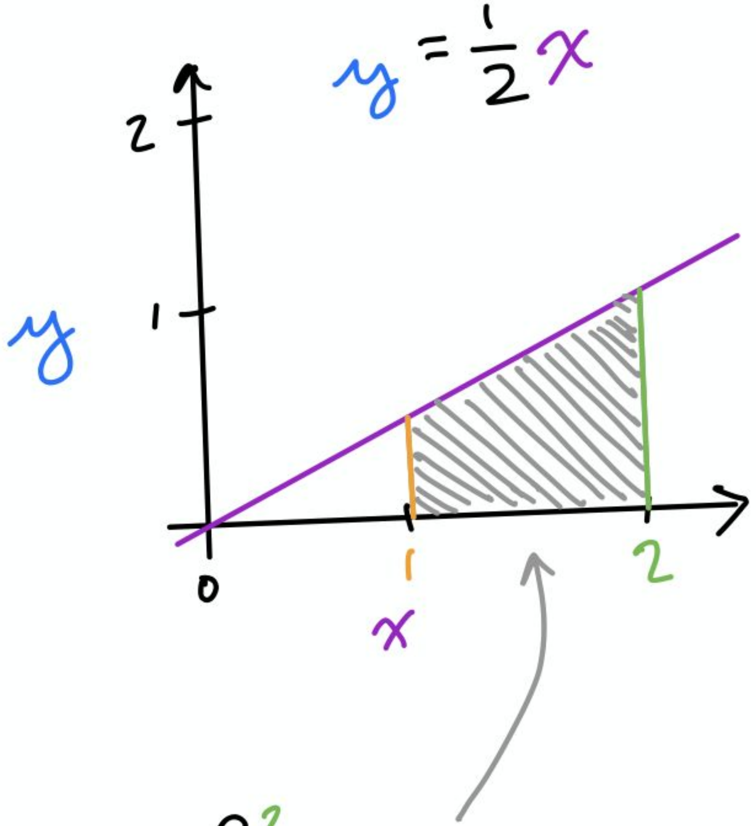




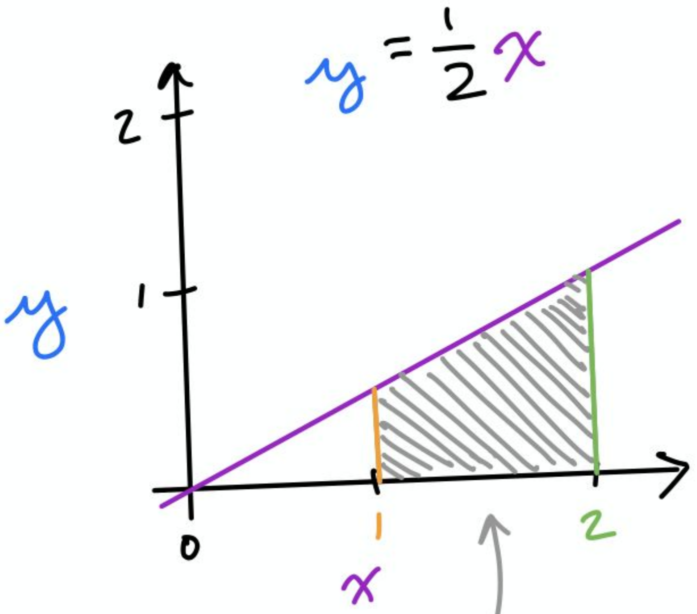
$y = \frac{1}{2}x$
y
x
0
1
2
1
2



y = 1/2 x
2
1
y
0
1
2
x

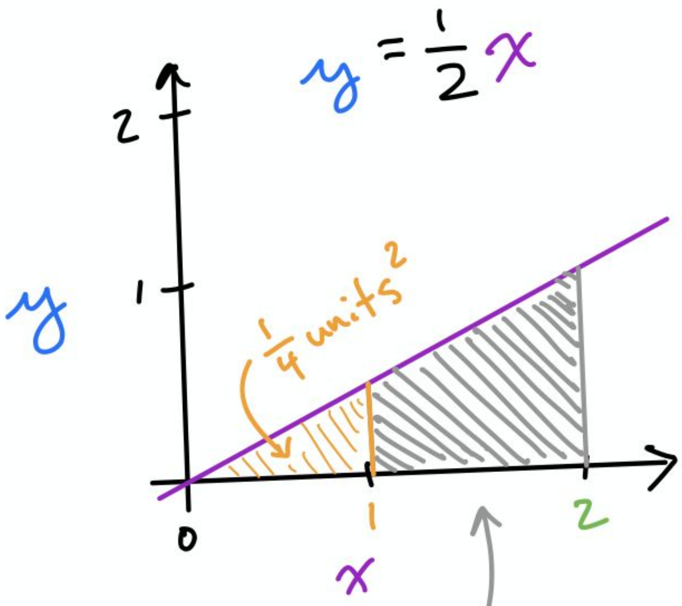


$$\int_1^2 \frac{1}{2}x \, dx$$



$$\int \frac{1}{2} x \, dx = \frac{1}{2}\left(\frac{x^2}{2}\right) + C = \frac{x^2}{4} + C$$

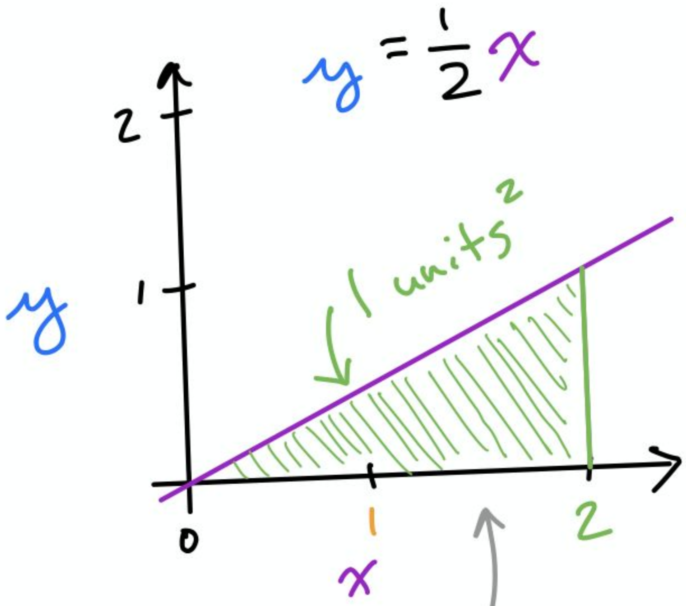
$$\int_1^2 \frac{1}{2} x \, dx$$



$$\int \frac{1}{2} x \, dx = \frac{1}{2}\left(\frac{x^2}{2}\right) + C = \frac{x^2}{4} + C$$

$$\boxed{\text{At } x = 1, \; \frac{x^2}{4} + C = \frac{1^2}{4} + C = \frac{1}{4} + C}$$

$$\int_1^2 \frac{1}{2} x \, dx$$

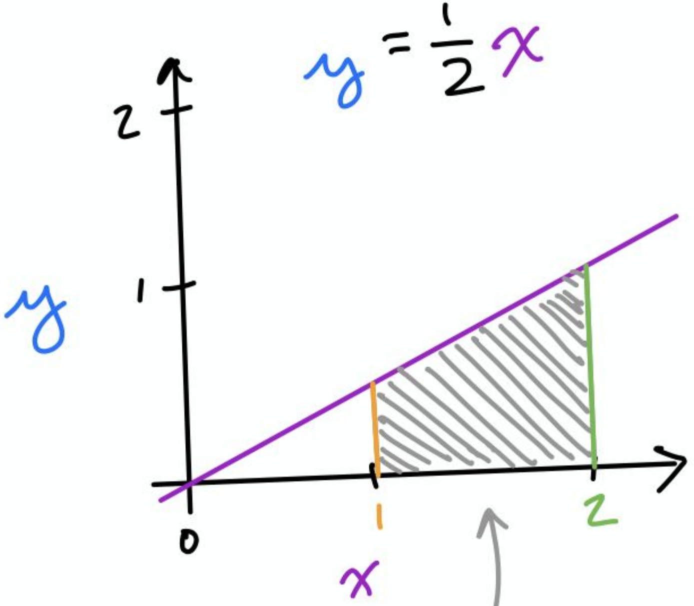


$$\int_{1}^{2} \frac{1}{2} x \, dx$$

$$\int \frac{1}{2} x \, dx = \frac{1}{2}\left(\frac{x^2}{2}\right) + C = \frac{x^2}{4} + C$$

At $x = 1$, $\frac{x^2}{4} + C = \frac{1^2}{4} + C = \frac{1}{4} + C$

$$\boxed{\text{At } x = 2,\ \frac{x^2}{4} + C = \frac{2^2}{4} + C = 1 + C}$$

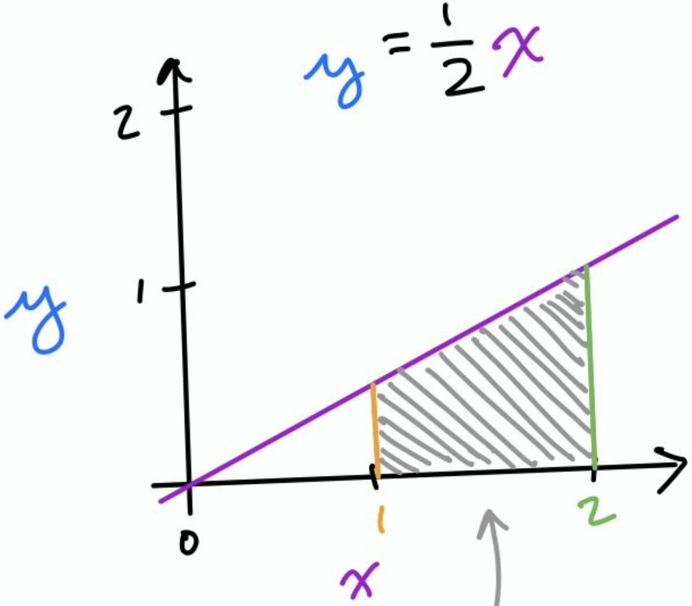


$$\int \frac{1}{2} x \, dx = \frac{1}{2}\left(\frac{x^2}{2}\right) + C = \frac{x^2}{4} + C$$

At $x = 1$, $\frac{x^2}{4} + C = \frac{1^2}{4} + C = \frac{1}{4} + C$

At $x = 2$, $\frac{x^2}{4} + C = \frac{2^2}{4} + C = 1 + C$

$$\int_1^2 \frac{1}{2} x \, dx = (1 + C) - \left(\frac{1}{4} + C\right)$$



$$\int \frac{1}{2}x\,dx = \frac{1}{2}\left(\frac{x^2}{2}\right) + C = \frac{x^2}{4} + C$$

$$\text{At } x = 1,\ \frac{x^2}{4} + C = \frac{1^2}{4} + C = \frac{1}{4} + C$$

$$\text{At } x = 2,\ \frac{x^2}{4} + C = \frac{2^2}{4} + C = 1 + C$$

$$\int_1^2 \frac{1}{2}x\,dx = (1 + C) - \left(\frac{1}{4} + C\right) = \frac{3}{4}$$

Exercise

Evaluate the following expression using both pencil and Python:

$$\int_3^4 2x \, dx$$

Solution

$$\int_3^4 2x\,dx$$

$$\int 2x\,dx = 2\left(\frac{x^2}{2}\right) + C = x^2 + C$$

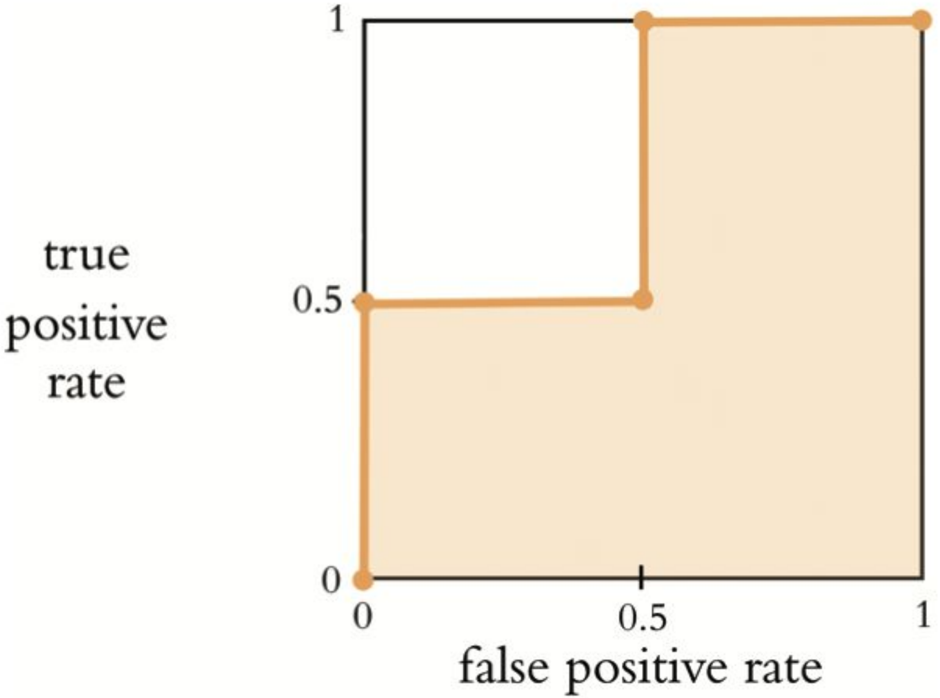
At $x = 3$, $x^2 + C = (3)^2 + C = 9 + C$

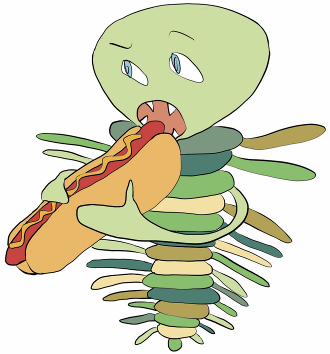
At $x = 4$, $x^2 + C = (4)^2 + C = 16 + C$

$$\int_3^4 2x\,dx = (16 + C) - (9 + C) = 7$$

Hands-on code demo

Area Under the ROC Curve





Hands-on code demo

Resources for Further Study

- General reference textbook: Michael Spivak's *Calculus*
- **Differential Calculus**:
  - Ch. 6 of Deisenroth et al. (2020) *Mathematics for ML*
  - 3Blue1Brown on YouTube
- **Integral Calculus**:
  - ditto
  - Appendix 18.5 of Zhang et al.'s (2019) *Dive into Deep Learning*
- **Next steps in the *ML Foundations* series:**
  - Probability & Information Theory
  - Intro to Statistics
  - Optimization

Next Subject: *Probability & Info. Thy.*

Apply calculus to ascertain how much meaningful signal is present in data.

Learn the probability theory-based foundations of stats and ML.

POLL *with Multiple Answers Possible*

What follow-up topics interest you most?

- Linear Algebra
- More Calculus
- Probability / Statistics
- Computer Science (e.g., algorithms, data structures)
- Machine Learning Basics
- Advanced Machine Learning, incl. Deep Learning
- Something Else

Stay in Touch

`jonkrohn.com` to sign up for email newsletter

`linkedin.com/in/jonkrohn`

`youtube.com/c/JonKrohnLearns`

`twitter.com/JonKrohnLearns`

NEBULA

PLACEHOLDER
FOR:

5-Minute Timer

PLACEHOLDER
FOR:

10-Minute Timer

PLACEHOLDER
FOR:

15-Minute Timer