Mathematics for Artificial Intelligence

Yogesh Kulkarni

January 22, 2021

About Content

- Bare essential mathematics needed for Machine/Deep Learning
- Just to get you started.
- Each field within Machine/Deep Learning can go extremely deep.
- This course is to give you enough foundation needed.

About Me

- I am not a mathematician
- And this course won't turn you into one!!! (if you are not already)
- But, I will surely attempt to make you get interested in learning more.

Calculus

- Algebra deals with finite processes.
- Calculus deals with infinitesimal processes.
- Limiting situations.
- Computers can not handle, need to approximate.

Main parts of Calculus

- Numbers: how the number-line ('x' axis for single variable) is constructed. Various sets with infinite items.
- Functions: Functions that generate numbers, their types.
- Most real-life sets are finite, e.g. number of people in the world, number of hair on head, etc.
- But numbers are infinite.

Basic Idea

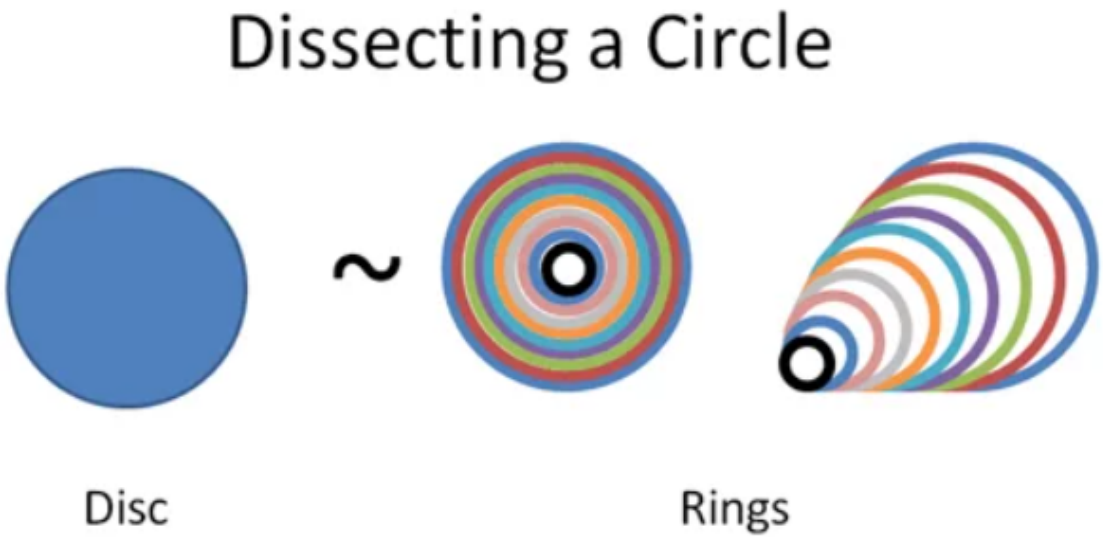
- Approximate/Divide
- Compose the total
- Reduce approximation infinitesimally small
- Get the actual results

Basic Idea: Example

- We know, circumference of circle is $2\pi r$
- Whats the equation of area?

Basic Idea: Example

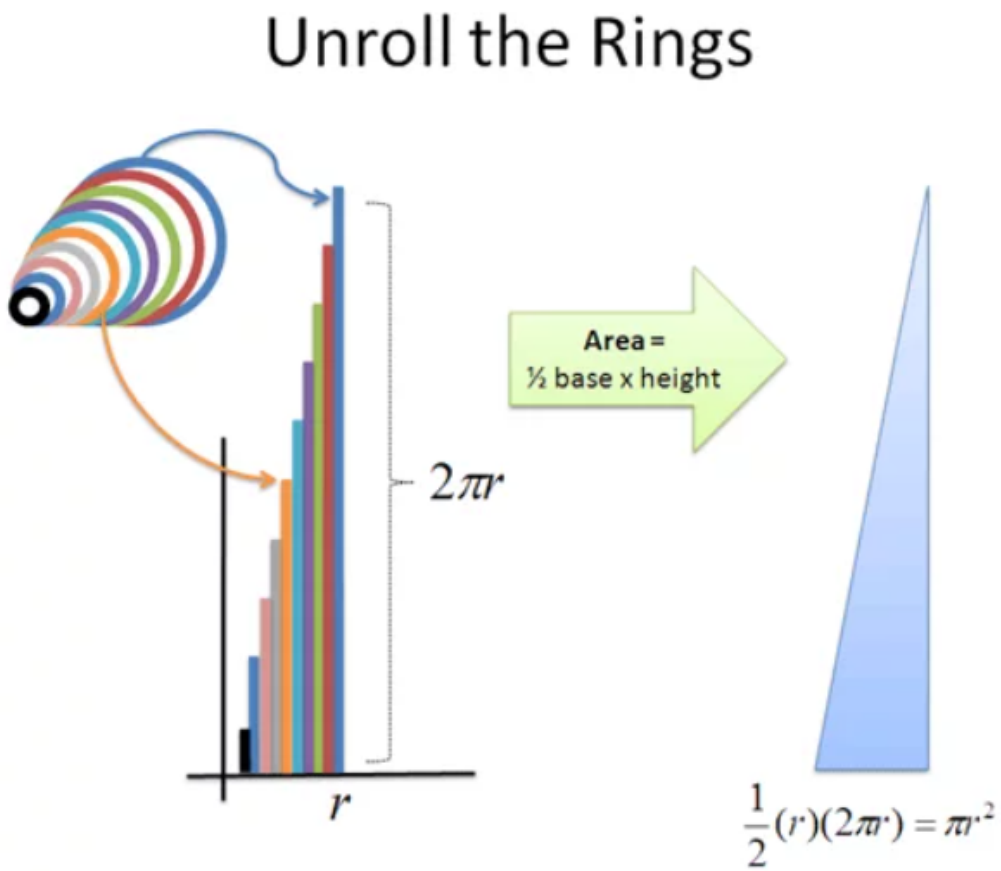
Divide the circle into concentric strips, of *small* widths.



(Ref: Lesson 1: 1 Minute Calculus: X-Ray and Time-Lapse Vision - Better Explained)

Basic Idea: Example

Compose to get the desired Area



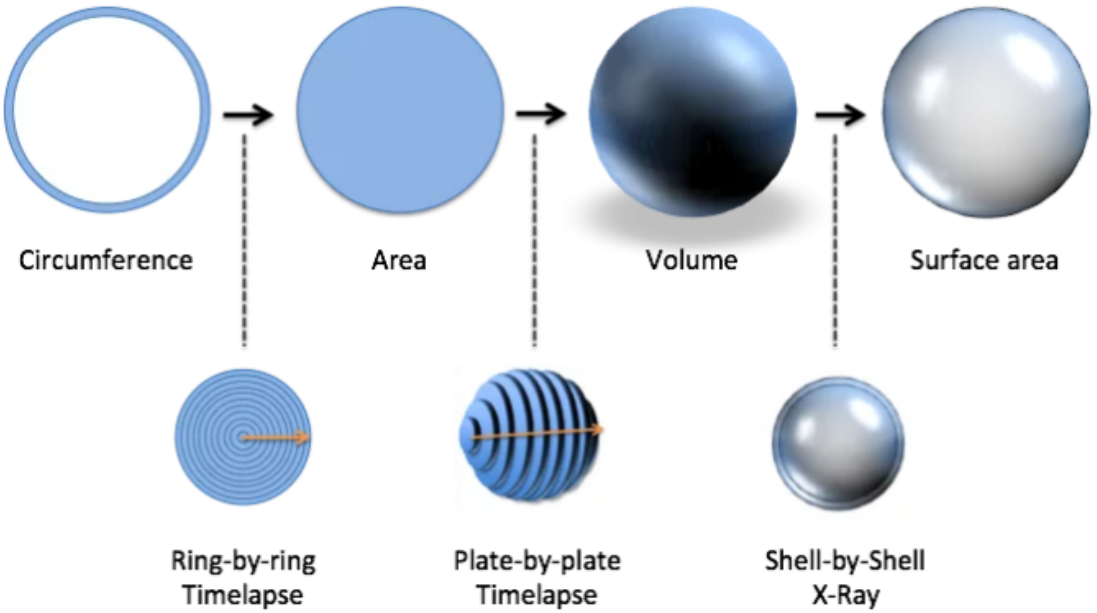
(Ref: Lesson 1: 1 Minute Calculus: X-Ray and Time-Lapse Vision - Better Explained)

Basic Idea: Example

- Still approximate, right?
- Reduce the approximation ie width to 'infinitesimally' small to get accurate results.
- This is called as . . . ??

(Ref: Lesson 1: 1 Minute Calculus: X-Ray and Time-Lapse Vision - Better Explained)

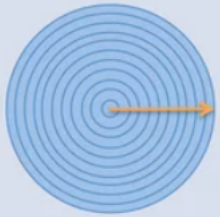
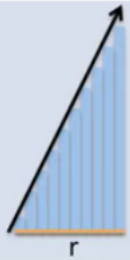
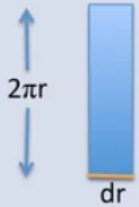
Basic Idea: Example



- Circumference with widths giving Area
- Area slices with widths giving Volume
- Slice places widths composed to give Surface Area

(Ref: Lesson 3: Expanding Our Intuition - Better Explained)

Basic Idea: Mathematical Notations

Strategy	Visualization	Step-by-Step Layout	Single Step Zoom
Ring-by-ring Timelapse			

(Ref: Lesson 4: Learning The Official Terms - Better Explained)

Basic Idea: Mathematical Notations

Intuitive Concept	Formal Name	Symbol
X-Ray (split apart)	Take the derivative (derive)	$\frac{d}{dr}$
Time-lapse (glue together)	Take the integral (integrate)	$\int$
Arrow direction	Integrate or derive "with respect to" a variable.	*dr* implies moving along *r*
Arrow start/stop	Bounds or range of integration	$\int_{start}^{end}$
Slice	Integrand (shape being glued together, such as a ring)	Equation, such as $2\pi r$

(Ref: Lesson 4: Learning The Official Terms - Better Explained)

Basic Idea: Algebra vs Calculus

Operation	Example	Notes
Division	$\frac{y}{x}$	Split whole into identical parts
Differentiation	$\frac{d}{dx}y$	Split whole into (possibly different) parts
Multiplication	$y \cdot x$	Accumulate identical steps
Integration	$\int y\,dx$	Accumulate (possibly different) steps

(Ref: Lesson 6: Improving Arithmetic And Algebra - Better Explained)

Basic Idea: More Formulas

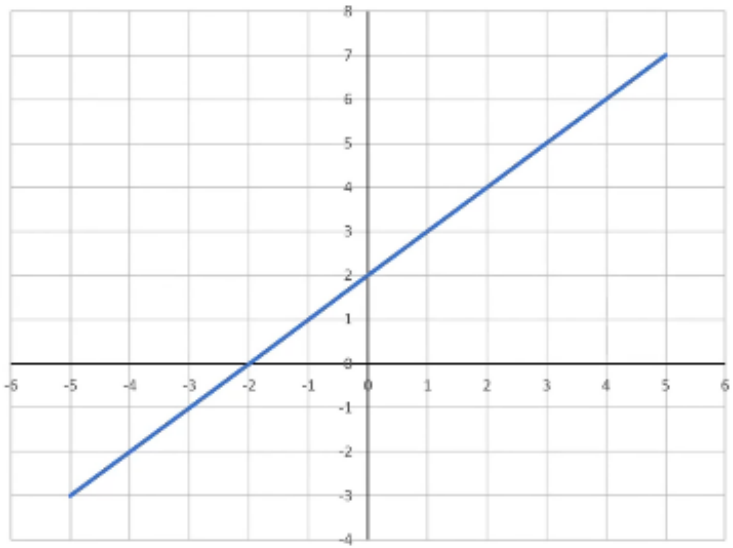
Algebra	Calculus
$distance = speed \cdot time$	$distance = \int speed\ dt$
$speed = \frac{distance}{time}$	$speed = \frac{d}{dt} distance$
$area = height \cdot width$	$area = \int height\ dw$
$weight = density \cdot length \cdot width \cdot height$	$weight = \iiint density\ dx\ dy\ dz$

(Ref: Lesson 6: Improving Arithmetic And Algebra - Better Explained)

Functions

Functions

- Takes many inputs and produces one output
- $f(x) = x + 2$
- $f(3) = ?$
- This function is a line, with y is same as $f(x)$



Functions

- Another example : $g(x) = 10/x$
- $g(5) =?$
- $g(0) = \textit{undefined}$
- Set of numbers for which the function is defined is called "Domain".
- For g domain is $\{x \in \mathbb{R} | x \neq 0\}$

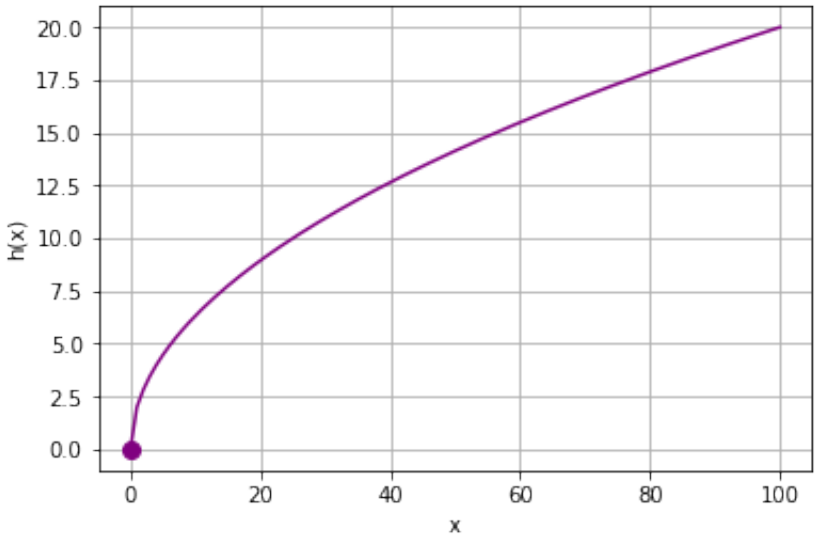
Exercise

$h(x) = 2\sqrt{x}, x \geq 0$

```
1  import numpy as np
   from matplotlib import pyplot as plt
3
   def h(x):
5      if x >= 0:
           return 2 * np.sqrt(x)
7
   x = range(-100, 101)
9  y = [h(a) for a in x]

11 plt.xlabel('x')
   plt.ylabel('h(x)')
13 plt.grid()

15 plt.plot(x,y, color='purple')
   plt.plot(0,h(0.01), color='purple', marker='o')
17 plt.show()
```

See, although 'domain' is infinite, for plotting, we need to restrict it to certain range.

Plot



Limits

Limits, the Foundations Of Calculus

What's this? . . .

Let x approach 0, but not get there, yet we'll act like it's there . . .

What's going on . . .

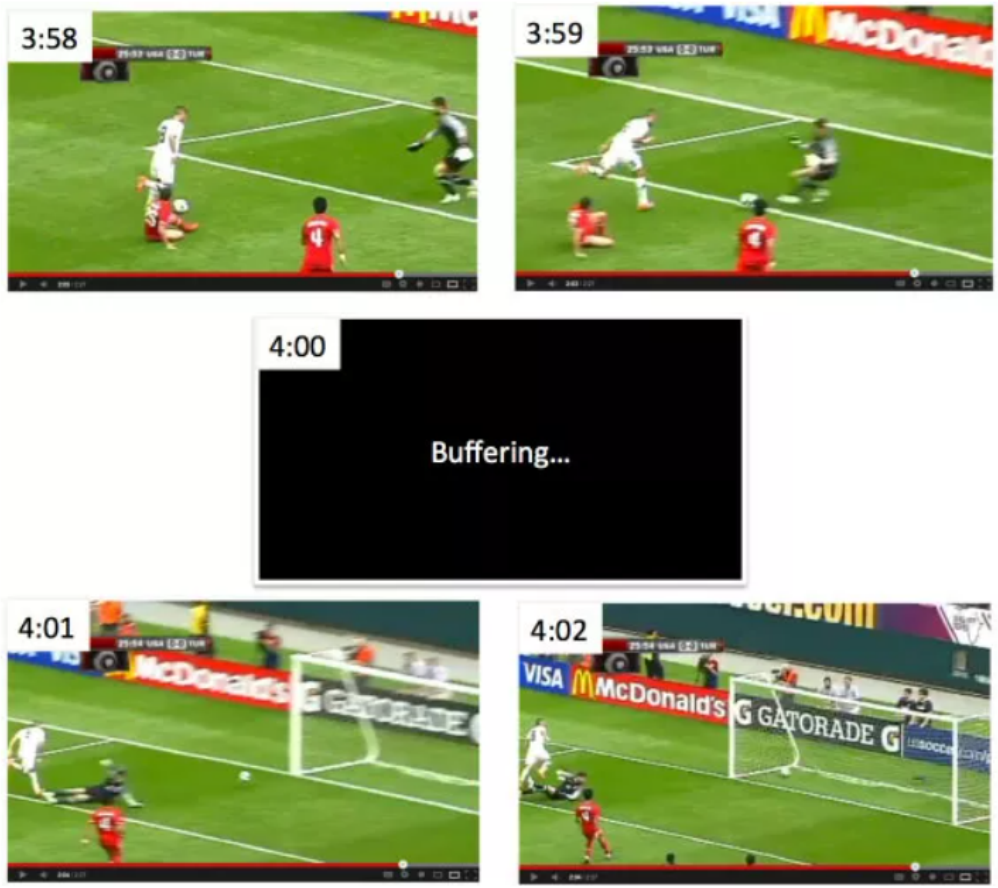
(Ref: An Intuitive Introduction To Limits - Better Explained)

Learnings

- **What is a limit?**: Our best **prediction** of a point we didn't observe.
- **How do we make a prediction?**: Zoom into the neighboring points. If our prediction is always in-between neighboring points, no matter how much we zoom, that's our estimate.
- **Why do we need limits?**: Math has "black hole" scenarios (dividing by zero, going to infinity), and limits give us an estimate when we can't compute a result directly.
- **How do we know we're right?**: We don't. Our prediction, the limit, isn't required to match reality. But for most natural phenomena, it seems to.
- **Limits let us ask "What if?"**: If we can directly observe a function at a value (like x=0, or x growing infinitely), we don't need a prediction. The limit wonders, "If you can see everything except a single value, what do you think is there?".
- **Prediction**: When our prediction is consistent and improves the closer we look, we feel confident in it. And if the function behaves smoothly, like most real-world functions do, the limit is where the missing point must be.

(Ref: An Intuitive Introduction To Limits - Better Explained)

Say, you missed a frame



Can you guess/predict ball position at the missing frame?

(Ref: An Intuitive Introduction To Limits - Better Explained)

Are you sure?

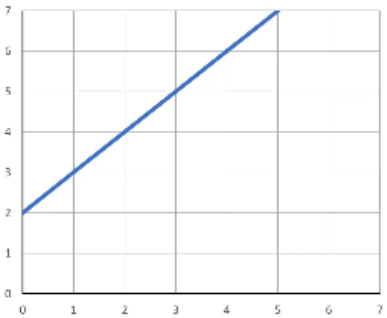
- Theoretically, the ball could have been anywhere, right?
- We just don't know, but we can predict.
- The predictions agree at increasing zoom levels.
- The before-and-after agree.
- On the contrary: Imagine at 3:59 the ball was at 10 meters, rolling right, and at 4:01 it was at 50 meters, rolling left. What happened? We had a sudden jump (a camera change?) and now we can't pin down the ball's position. Which one had the ball at 4:00? This ambiguity shatters our ability to make a confident prediction.

(Ref: An Intuitive Introduction To Limits - Better Explained)

Limits

- Functions can be plotted, with input(s) on say x and output on y (higher dimension functions, not considered as of now)
- Say, $f(x) = x + 2$ is function showing distance traveled (y) at each x seconds.
- Slope gives rate of change, ie Speed.and its constant.
- It can be calculated by taking any two points on the line

$$f(x) = x + 2$$



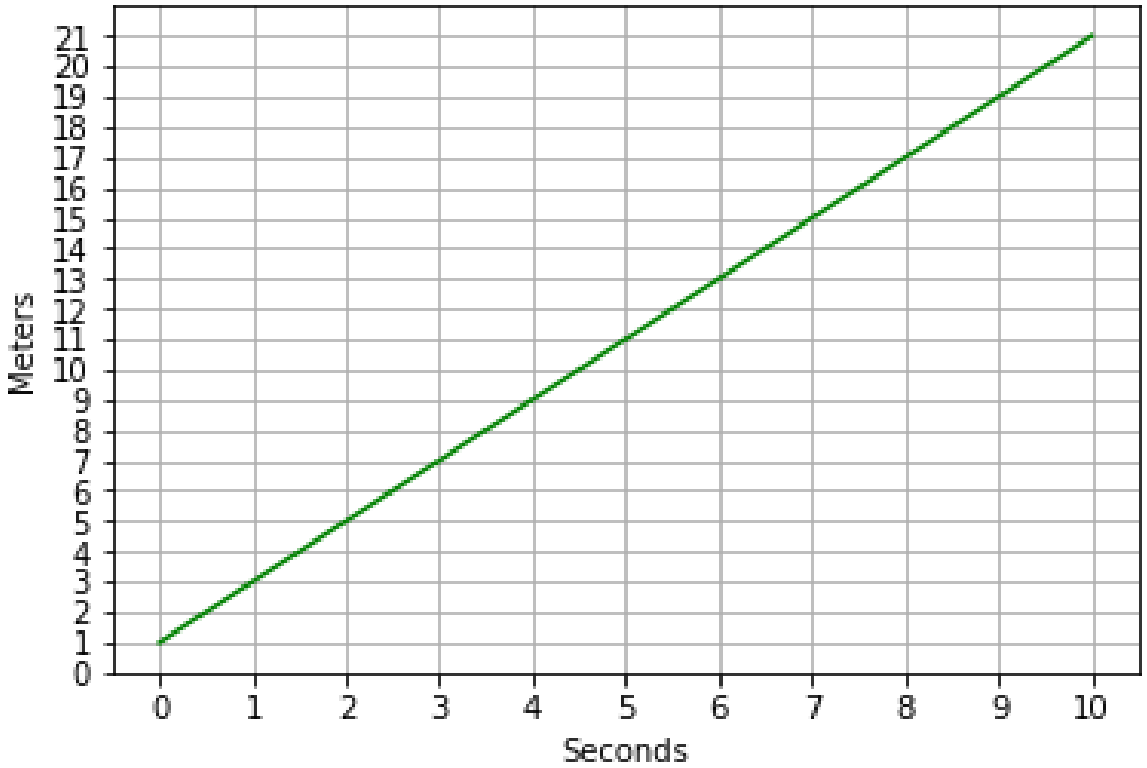
$$m = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

Exercise

$$q(x) = 2x + 1$$

```
import numpy as np
from matplotlib import pyplot as plt

def q(x):
    return 2*x + 1

x = np.array(range(0, 11))

plt.xlabel('Seconds')
plt.ylabel('Meters')
plt.xticks(range(0,11, 1))
plt.yticks(range(0, 22, 1))
plt.grid()

plt.plot(x,q(x), color='green')
plt.show()
```

Plot

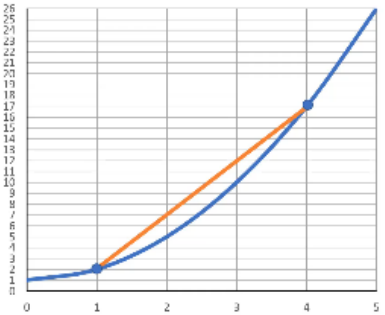


Is Slope same/different on any point on the line?

Functions

- If $f(x) = x^2 + 1$ is function showing distance traveled (y) at each x seconds.
- Secant Slope gives rate of change, ie Speed. and its approximate.
- For different secants, different slopes, so any change in speed is called acceleration.

$$f(x) = x^2 + 1$$



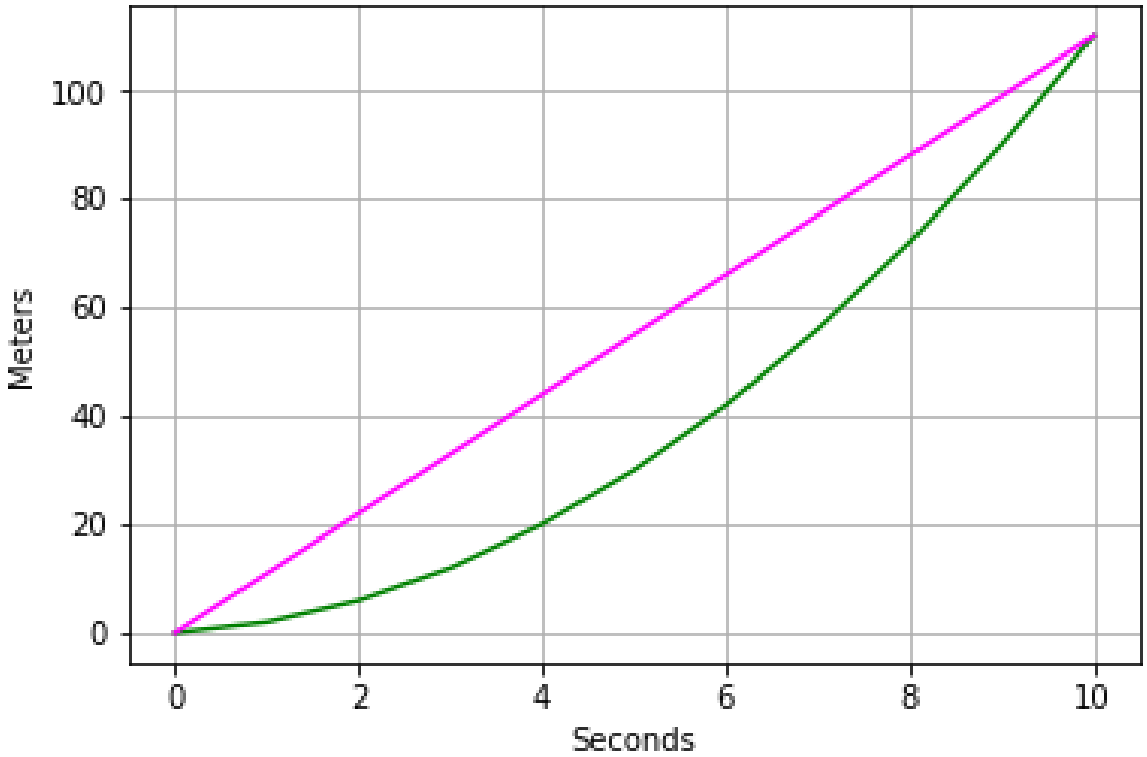
$$m = \frac{17 - 2}{4 - 1}$$

Exercise

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

```
import numpy as np
from matplotlib import pyplot as plt

def r(x):
    return x**2 + x

x = np.array(range(0, 11))
s = np.array([0,10])

plt.xlabel('Seconds')
plt.ylabel('Meters')
plt.grid()

# Plot x against r(x)
plt.plot(x,r(x), color='green')

# Plot the secant line
plt.plot(s,r(s), color='magenta')
plt.show()
```

Plot



Average velocity is 10 m/s

Practical Example

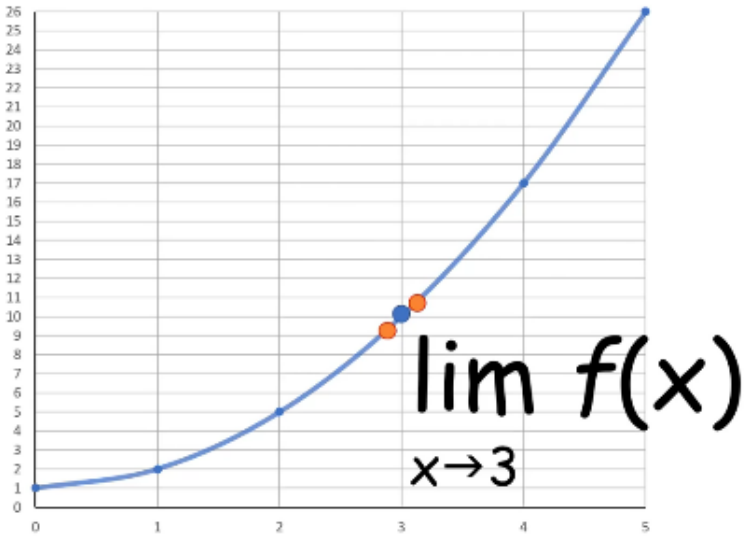
- Say, Pune (P), and Mumbai (M) are 200km apart.
- It takes 4hrs to cover the distance, so the average velocity is $\frac{200}{4} = 50 kmph$.
- Is that the velocity throughout the journey? No. We want more refined answer.
- At Wakad, it could be 20; on highway it could be 100, etc.
- Say, you want to know: within Lonawala, which is about 1km wide-long, whats the velocity?: you can measure from start, end of 1km, and decide.
- Going narrower: What's the velocity while crossing Maganlal shop? which is about 10m long?
- Narrowing down more and more, to find what is called 'instantaneous' velocity.
- Very small, infinitesimally small, a point.
- Thats the "Limiting condition"

Instantaneous Velocity?

- Secant is an approximation, not an exact speed at a particular moment.
- We need slope of a curve AT A SINGLE POINT (and not approximate average between two secant points)
- If we want slope at a particular x, we can find slope of a secant between x1 and x2 which are very close to x.

Instantaneous Velocity?

- For a point on the curve ,say, (3,10)
- From left and right sides we can approach the point.
- Y value at such nearby points is called as LIMIT
- And its written as:



The Formal Definition Of A Limit

$$\lim_{x \to c} f(x) = L$$

means for all real $\varepsilon > 0$ *there exists a real* $\delta > 0$ *such that for all x with* $0 < |x - c| < \delta$, *we have* $|f(x) - L| < \varepsilon$

Math English	Human English
$\lim_{x \to c} f(x) = L$ means	When we "strongly predict" that f(c) = L, we mean
for all real $\varepsilon > 0$	for any error margin we want (+/- .1 meters)
there exists a real $\delta > 0$	there is a zoom level (+/- .1 seconds)
such that for all x with $0 < \|x - c\| < \delta$, we have $\|f(x) - L\| < \varepsilon$	where the prediction stays accurate to within the error margin

(Ref: An Intuitive Introduction To Limits - Better Explained)

A few subtleties

$$\lim_{x \to c} f(x) = L$$

means for all real $\varepsilon > 0$ *there exists a real* $\delta > 0$ *such that for all* x *with* $0 < |x - c| < \delta$, *we have* $|f(x) - L| < \varepsilon$

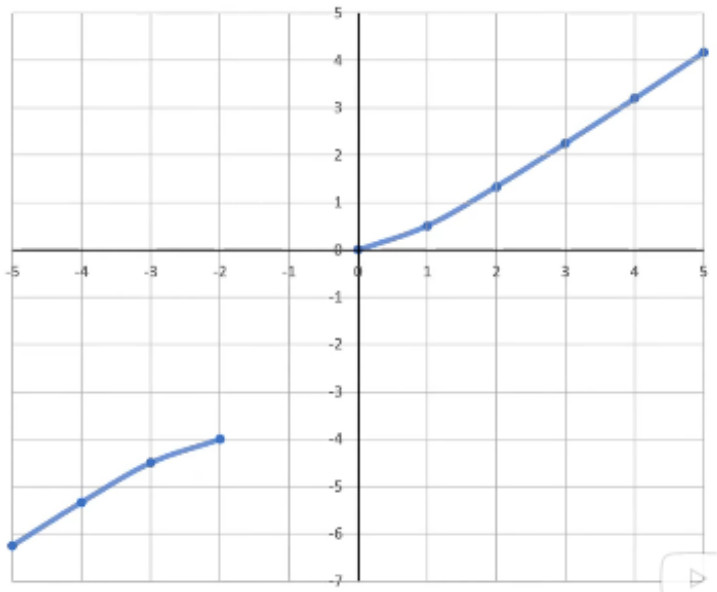
- The zoom level (delta, δ) is the function input, i.e. the time in the video
- The error margin (epsilon, ϵ) is the most the function output (the ball's position) can differ from our prediction throughout the entire zoom level
- The absolute value condition ($0 < |x - c| < \delta$) means positive and negative offsets must work, and we're skipping the black hole itself (when $|x - c| = 0$).

$$\lim_{x \to c} f(x) = f(c)$$

(Ref: An Intuitive Introduction To Limits - Better Explained)

Continuity

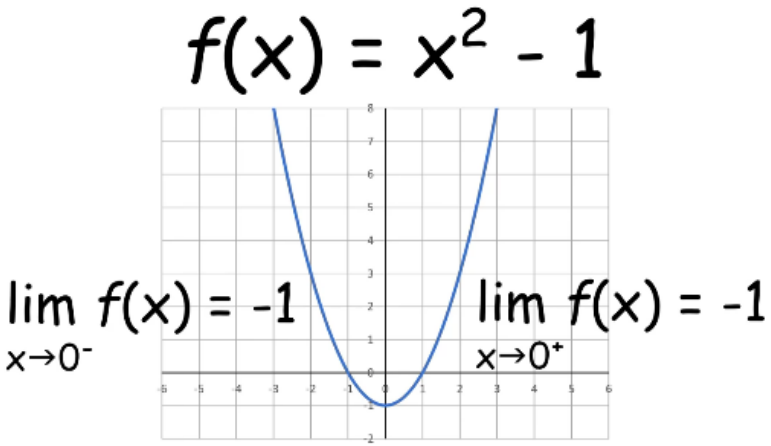
- Most functions we looked at so far, are continuous for all real values of x
- Meaning if you trace them, you need not lift your pen.
- Not all functions are defined like that.
- $f(x) = \frac{x^2}{x+1}, x \leq -2 | x \geq 0$
- Obviously, there is a gap in the domain as well as output, so not continuous.



Finding Limits

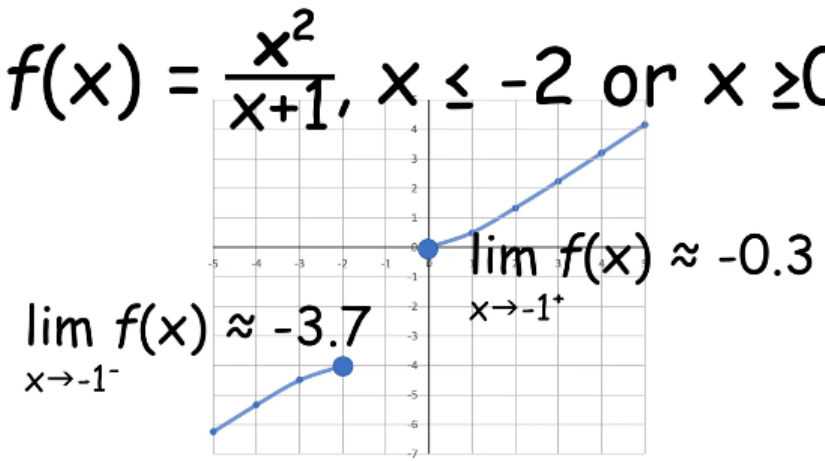
What is meant by finding limits of a function at a particular x?

- Go from left side (ie less of x, increasing) and see to which value $f(x)$ is approaching.
- Go from right side (ie more of x, decreasing) and see to which value $f(x)$ is approaching.
- If both results are same, we got the limit value.
- Even if $f(x)$ is not defined for that x, the limit value exists!!!



Finding Limits

- Lets look at obvious non continuous function
- Need to find limit of function at x = -1
- From negative side, the $f(x)$ is projected to go to 3.7 or so for x = -1
- From positive side, the $f(x)$ is projected to go to -0.3 so for x = -1
- Both do NOT agree. So limit does not exists.



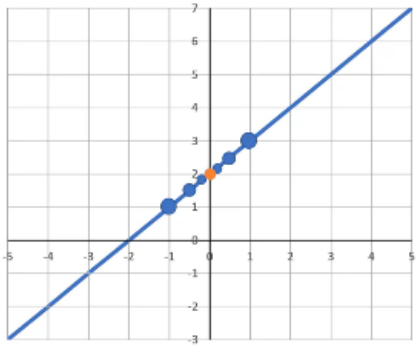
Finding Limits

- Lets look at single point non continuous function
- Need to find limit of function at x = 0
- Calculate values from both sides
- The limit is approaching to 2

$$f(x) = x + 2, x \neq 0$$

x	f(x)
-1	1
-0.5	1.5
-0.01	1.99
0.01	2.01
0.5	2.5
1	3

2



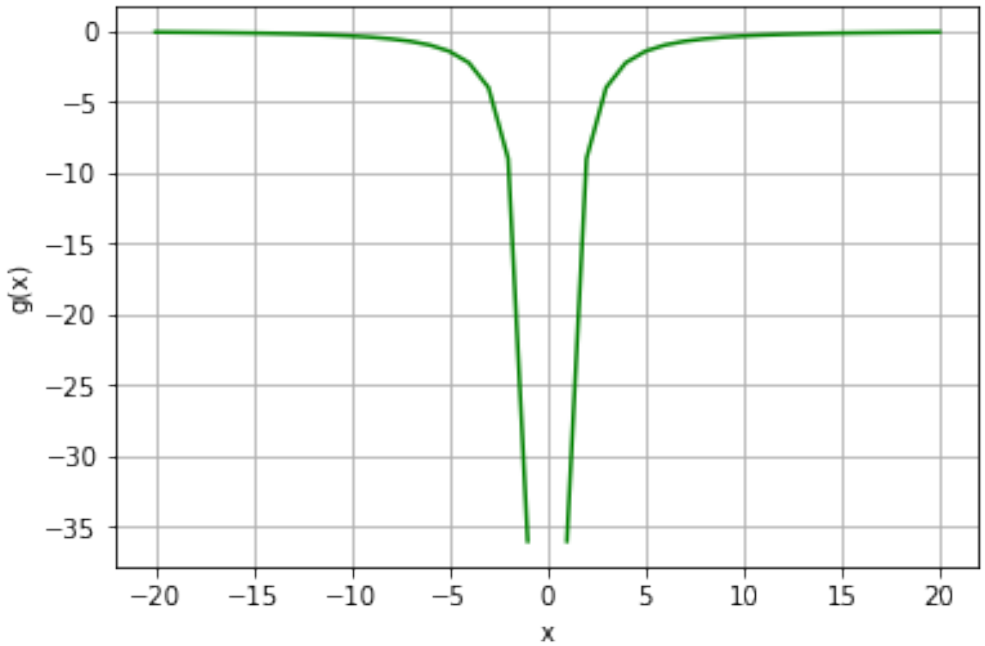
Why not just substitute the value?

Exercise

$$g(x) = -(\frac{12}{2x})^2, x \neq 0$$

```
from matplotlib import pyplot as plt

def g(x):
    if x != 0:
        return -(12/(2*x))**2

x = range(-20, 21)
y = [g(a) for a in x]

plt.xlabel('x')
plt.ylabel('g(x)')
plt.grid()

# Plot x against g(x)
plt.plot(x,y, color='green')

plt.show()
```

Plot



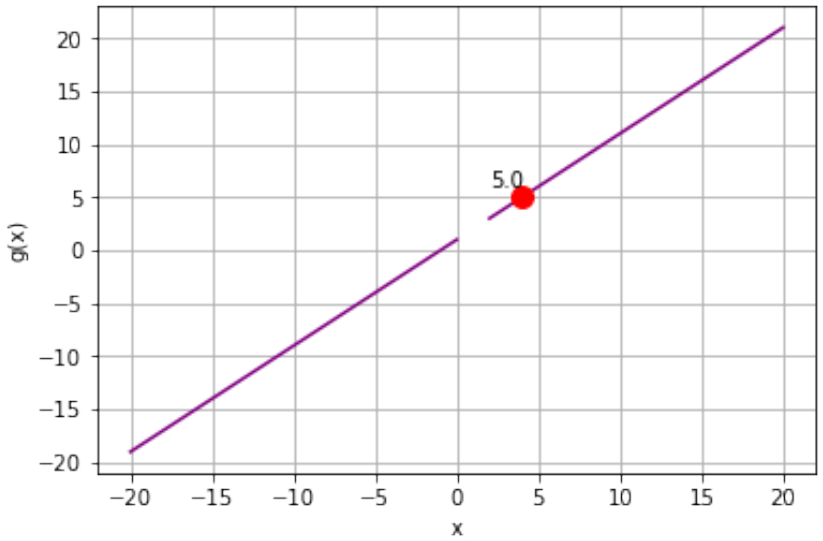
Exercise

$d(x) = \frac{4}{x-25}, x \neq 25$

```
from matplotlib import pyplot as plt

def d(x):
    if x != 25:
        return 4 / (x - 25)

x = list(range(-100, 24))
x.append(24.9) # Add some fractional x
x.append(25)   # values around
x.append(25.1) # 25 for finer-grain results
x = x + list(range(26, 101))
# Get the corresponding y values from the function
y = [d(i) for i in x]

plt.xlabel('x')
plt.ylabel('d(x)')
plt.grid()

plt.plot(x,y, color='purple')
plt.show()
```

Finding Limits

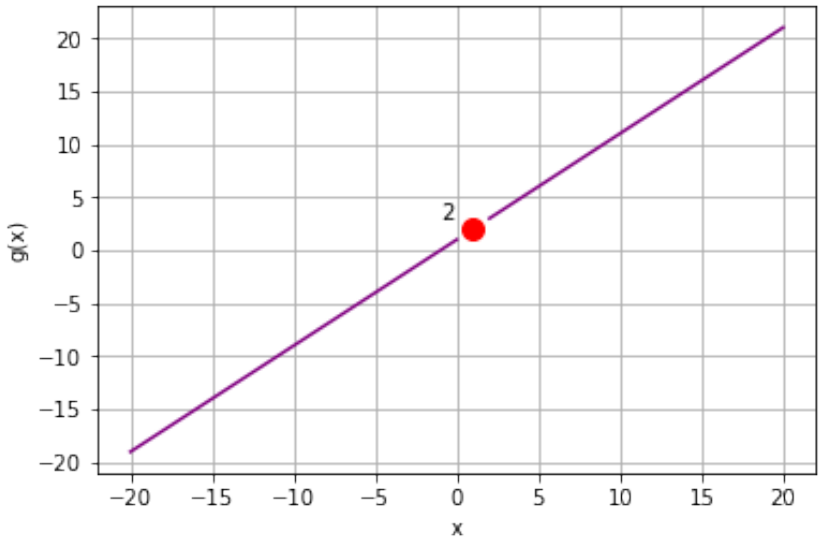
Direct Substitution $g(x) = \frac{x^2-1}{x-1}, x \neq 1$



Find the limit of g(x) as x approaches 4. Just substitute and get the limit as 5.

Finding Limits

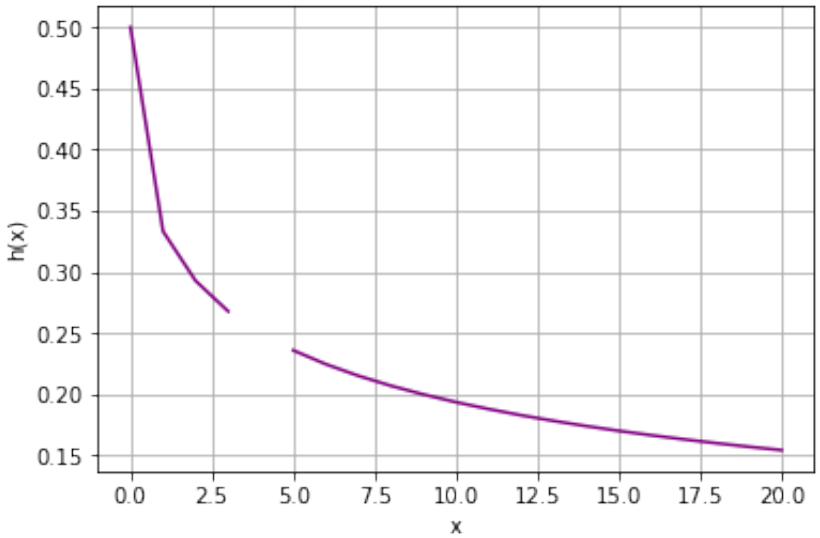
Factorization $g(x) = \frac{x^2-1}{x-1}, x \neq 1$



Find the limit of g(x) as x approaches 1. Substituting gives 0/0, so not good. Factorize and cancel common portion $(x-1)$. Then substitute. Result is 2.

Finding Limits

Rationalization $h(x) = \frac{\sqrt{x}-2}{x-4}, x \neq 4 and x \geq 0$



Multiply top and bottom by $\sqrt{x}+2$. Then substitute x = 4.. Result is 0.25

Rules of Limits

- $\lim_{x \to a}(j(x) + l(x)) = \lim_{x \to a} j(x) + \lim_{x \to a} l(x)$
- $\lim_{x \to a}(j(x) - l(x)) = \lim_{x \to a} j(x) - \lim_{x \to a} l(x)$
- $\lim_{x \to a}(j(x) \cdot l(x)) = \lim_{x \to a} j(x) \cdot \lim_{x \to a} l(x)$
- $\lim_{x \to a} \frac{j(x)}{l(x)} = \frac{\lim_{x \to a} j(x)}{\lim_{x \to a} l(x)}$
- $\lim_{x \to a}(j(x))^n = \left(\lim_{x \to a} j(x) \right)^n$

Quiz

True or False:
If a function f is not defined at $x = a$ then the limit $lim_{x \to a} f(x)$ never exists.

Answer

False.
$lim_{x \to a} f(x)$ may exist even if function f is undefined at $x = a$. The concept of limits has to do with the behavior of the function close to $x = a$ and not at $x = a$.

Quiz

True or False:
$lim_{x \to a} f(x) = +\infty$
and
$lim_{x \to a} g(x) = +\infty$
then
$lim_{x \to a}[f(x) - g(x)]$ is always equal to 0.

Answer

False.
Infinity is not a number and $\infty - \infty$ is not equal to 0. $+\infty$ is a symbol to represent large but undefined numbers. $-\infty$ is small but undefined number.

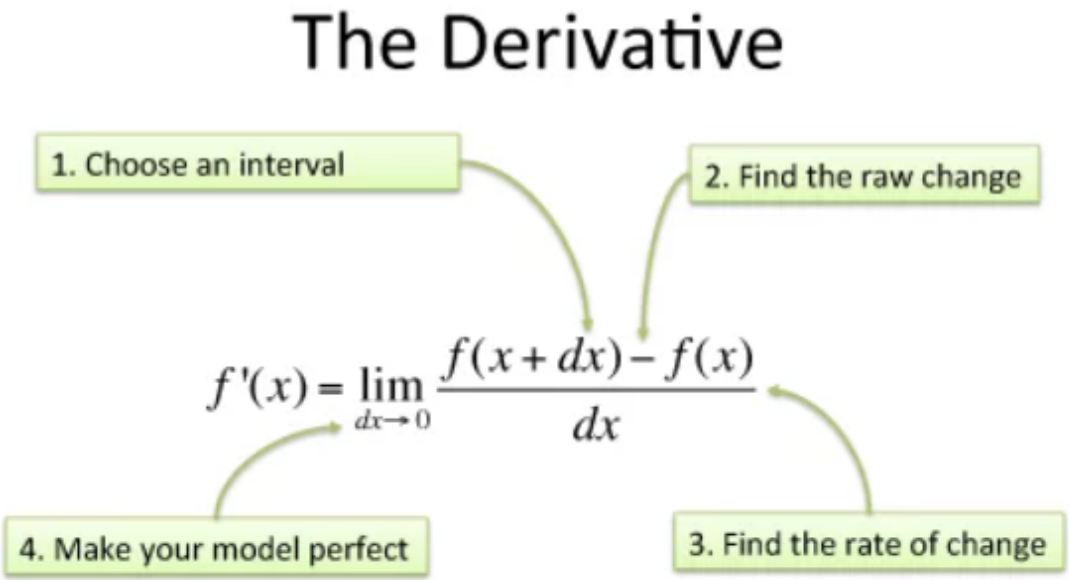
Differentiation

Background Theories

- **Limits**: We let the measurement artifacts get smaller and smaller until they effectively disappear (cannot be distinguished from zero).
- **Infinitesimals**: Create a new type of number that lets us try infinitely-small change on a separate, tiny number system. When we bring the result back to our regular number system, the artificial elements are removed.

(Ref: Lesson 10: The Theory Of Derivatives - Better Explained)

The Formula



(Ref: Lesson 10: The Theory Of Derivatives - Better Explained)

The Steps

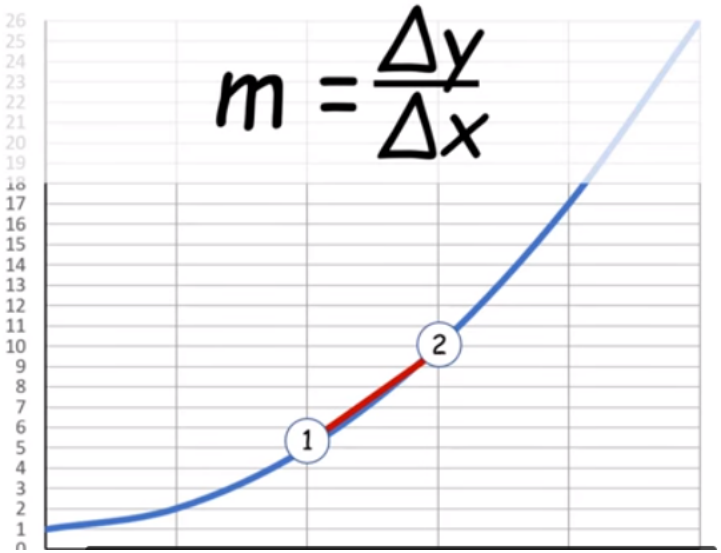
Step	Example
Start with function to study	$f(x) = x^2$
1. Increase the input by dx, a sample change	$f(x+dx) = (x+dx)^2 = x^2 + 2x \cdot dx + (dx)^2$
2. Find the resulting increase in output, df	$df = f(x+dx) - f(x) = 2x \cdot dx + (dx)^2$
3. Find the ratio of output change to input change	$\frac{df}{dx} = \frac{2x \cdot dx + (dx)^2}{dx} = 2x + dx$
4. Throw away any measurement artifacts	$2x + dx \overset{dx=0}{\Longrightarrow} 2x$

(Ref: Lesson 10: The Theory Of Derivatives - Better Explained)

Derivative

Slope of a graph between a section (ie two points) can be easily calculated.

$$f(x) = x^2 + 1$$

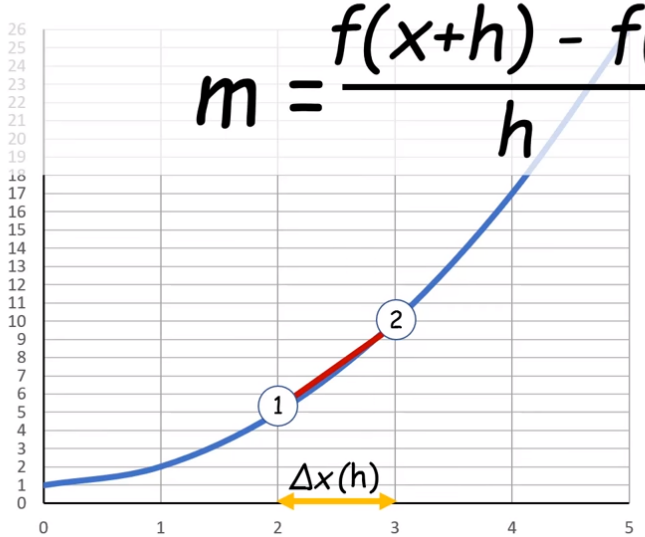


Derivative

Delta x can be replaced by h and the formula for segment slope becomes.

$$f(x) = x^2 + 1$$

$$m = \frac{f(x+h) - f(x)}{h}$$

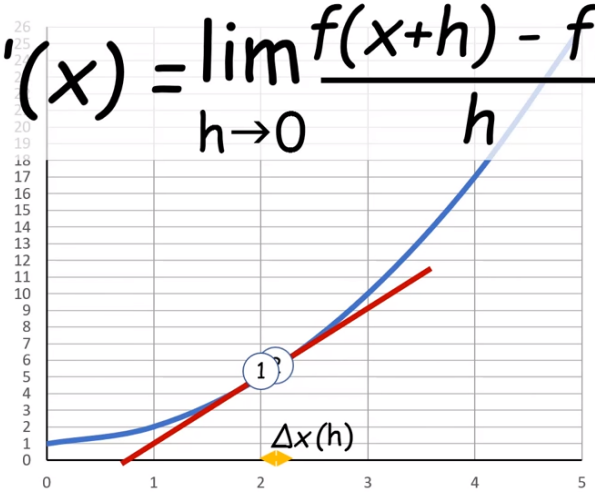


Derivative

If both points come very close then we can say that the 'Segment Slope' becomes a 'Point Slope'.

$$f(x) = x^2 + 1$$

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$



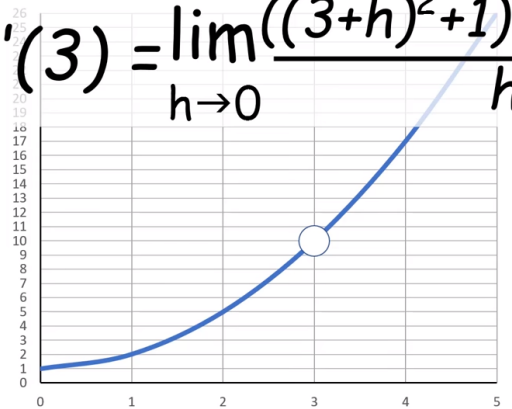
The concept of Limits is used to determine RATE OF CHANGE at a given POINT.

Derivative

How to calculate value of the derivative at a points, say $x = 3$?

$$f(x) = x^2 + 1$$

$$f'(3) = \lim_{h \to 0} \frac{((3+h)^2+1) - (3^2+1)}{h}$$



$f'(3) = 6$

Derivative

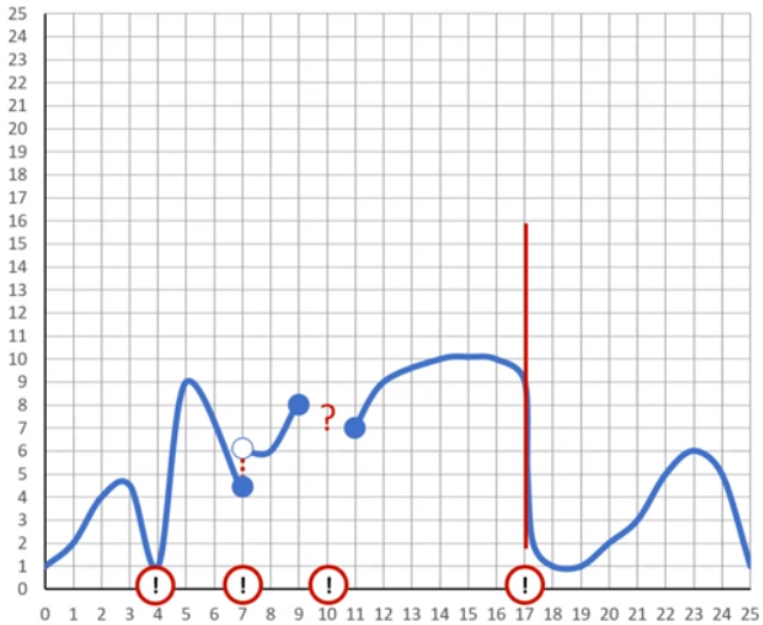
- Slope at a point on curve is linear, where as the curve itself is not linear.
- Tangent plane at a point on a sphere is linear but the sphere is not.
- So, derivative is a Linear approximation

Differentiability

- Note: A function may not be differentiable at every point; that is you might not be able to calculate the derivative for every point on the function line.
- To be differentiable at a given point, the function must be continuous at that point,
- The tangent line at that point cannot be vertical, and the line must be smooth at that point
- Cannot suddenly change direction or have kinks

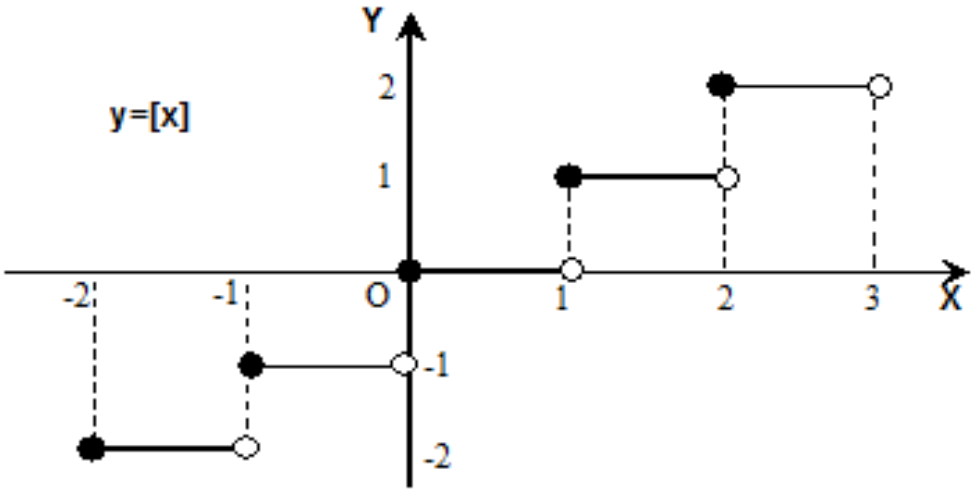
Continuous Functions

- A function below is not defined at certain points ($x = 10$)
- At certain point ($x = 7$), it is multi valued.
- At certain point ($x = 17$), the slope is vertical, so not differentiable.
- At certain point ($x = 4$), there is sudden change in direction, so not differentiable.



Discontinuous Functions

Step function: $f(x) = [x]$ integer part of x. This suddenly jumps from 0 to 1 to 2, etc.



Domain → R;
Period → non periodic;

Range → I;
Nature → neither even nor odd

Differentiability

- $f'(x_0) = \lim_{h \to 0} \frac{f(x_0+h)-f(x_0)}{h} = A$ if limit exists.
- In the above formula x_0 is constant, so this function is wrt h
- So, $\lim_{h \to 0} g(h) = A$
- For any given arbitrary $\epsilon > 0$, there exists $\delta > 0$, so that if $0 < |h| < \delta$, then $|g(h) - A| < \epsilon$
- However small ϵ can be made, we can still find corresponding δ. The dependence is on f as well as the point at which we are calculating.

Example

- $f(x) = x^2$, at $x_0 = 1$
- $f'(x) = \lim_{hto0} \frac{(x+h)^2 - x^2}{h} = 2x + h$
- Prove: $(x^n)' = nx^{n-1}$

$$(x^n)' = \lim_{h \to 0} \frac{(x+h)^n - x^n}{h} \tag{1}$$

$$= \lim_{h \to 0} \frac{x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 + \cdots + h^n - x^n}{h} \tag{2}$$

$$= \lim_{h \to 0} \left[nx^{n-1} + \frac{n(n-1)}{2}x^{n-2}h + \cdots + h^{n-1} \right] \tag{3}$$

- $\lim_{h \to 0} \left[nx^{n-1} + \frac{n(n-1)}{2}x^{n-2}h + \cdots + h^{n-1} \right] = nx^{n-1}$

Example

- $f(x) = x^2 + x$, at $x_0 = 5$
- $f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$
- $f'(x) = \lim_{h \to 0} \frac{((x+h)^2+x+h)-(x^2+x)}{h}$
- $f'(x) = \lim_{h \to 0} \frac{x^2+h^2+2xh+x+h-x^2-x}{h}$
- $f'(x) = \lim_{h \to 0} 2x + h + 1$
- Evaluating the limit, ie $h = 0$, becomes $f'(x) = 2x + 0 + 1$
- For $x = 5$ the result is $f'(5) = 2 \cdot 5 + 1 = 10 + 1 = 11$

Differentiation formulas

Let $f = f(x)$ and $g = g(x)$ have derivatives at $x = a$. Then

- $(f+g)'(a) = f'(a) + g'(a)$
- $(f-g)'(a) = f'(a) - g'(a)$
- $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$
- $\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{[g(a)]^2}$

In general, the derivative $f'(a)$ may be used by other notations:

$$Df(a) \qquad \frac{d}{dx}f(a) \qquad \frac{df}{dx}(a) \qquad \frac{df}{dx}\Big|_{x=a} \qquad Df\Big|_{x=a}$$

Theorems

Let f and g be differentiable on I, then so are the following functions. Moreover,

- $(f+g)' = f' + g'$
- $(f-g)' = f' - g'$
- $(fg)' = f'g + fg'$
- $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$

Concept	Key Analogy / Takeaway
Multiplication Rule $(f \cdot g)' = f \cdot dg + g \cdot df$	Grow a garden on two sides; ignore the corner. 

(Ref: Lesson 4: Learning The Official Terms - Better Explained)

The chain rule

Let $g = g(x)$ be a differentiable function at a, and $f = f(x)$ differentiable at $g(a)$. Then the composite function $f \circ g$ is differentiable at a, and

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

In Leibniz notation, let $y = f(u)$ and $u = g(x)$ be both differentiable, then

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

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

- Find $f'(x)$ if $f(x) = \sqrt{x^2+1}$.
- Differentiate (a) $y = \sin(x^2)$ (b) $y = \sin^2 x$.
- Differentiate $f'(x)$ if $f(x) = (x^3-1)^{100}$.
- Find $f'(x)$ if $f(x) = \dfrac{1}{\sqrt[3]{x^2+x+1}}$.
- Find the derivative of $g(t) = \left(\dfrac{t-2}{2t+1}\right)^9$.

  4[pt]
- Differentiate $y = (2x+1)^5(x^3-x+1)^4$.
- Find $f'(x)$ if $f(x) = \sin\Big(\cos(\tan x)\Big)$.
- Differentiate $y = \sqrt{\sec x^3}$.

2nd order Derivative

2nd order differentiation

- Finding rate of change of rate of change !!
- Velocity is rate of change of Distance
- Whats rate of change of Velocity?

Need 2nd order derivative in finding type of point where derivative goes 0, ie maxima or minima.

Maxima or Minima

$$f(x) = x^2$$
$$f'(x) = 2x$$
$$x = 0$$
$$f''(x) = 2$$

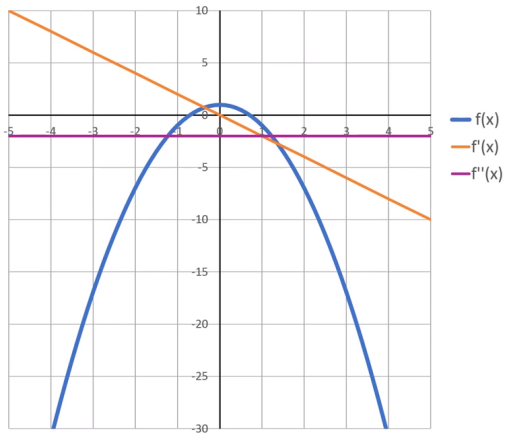
The derivative itself is a function. Its 2nd order derivative is coming to be a constant $+2$. So first order derivative is crossing from negative to positive. Thus its a Minima point.

Maxima or Minima

$f(x) = -x^2+1$

$f'(x) = -2x$

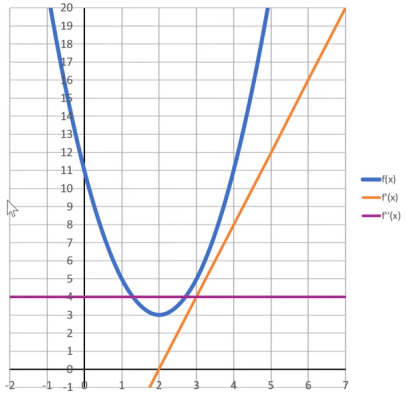
$f''(x) = -2$



Its 2nd order derivative is coming to be a constant -2. So first order derivative is crossing from positive to negative. Thus its a Maxima point.

Optimization

$$f(x) = 2(x-2)^2 + 3$$
$$= 2x^2 - 8x + 11$$
$$f'(x) = 4x - 8$$
$$x = 2$$
$$f''(2) = 4$$



First derivative goes to 0 at $x = 2$. At that point second derivative is positive. Thus its a Minima point.

Partial Derivative

Partial Derivative

- Functions can be of more-than-one-variable, say two variables: x and y.
- When plotted it will be 3D plot. 2 input variables and 1 output variable, total 3.
- For finding derivatives of such functions: we do it turn by turn.
- First wrt x then wrt y and so on. While taking derivative wrt x, the y is considered as constant.

$$f(x,y) = x^2 + y^2$$

$$\frac{d(x,y)}{dx} = \frac{d(x^2+y^2)}{dx} \quad \frac{d(x,y)}{dy} = \frac{d(x^2+y^2)}{dy}$$

$$\frac{dx^2}{dx} = 2x \quad \frac{dy^2}{dx} = 0 \qquad \frac{dx^2}{dy} = 0 \quad \frac{dy^2}{dy} = 2y$$

$$\frac{d(x,y)}{dx} = 2x \qquad \frac{d(x,y)}{dy} = 2y$$

Gradient

1. Partial derivatives are important if you want to find the analog of the slope for multi-dimensional surfaces. We call this quantity the gradient.
2. For $f(x,y) = x^2 + y^2$ the partial derivatives are:

$$\frac{\partial f(x,y)}{\partial x} = 2x \tag{4}$$

$$\frac{\partial f(x,y)}{\partial y} = 2y \tag{5}$$

3. The Gradient

$$grad(f(x,y)) = g(\vec{x,y}) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

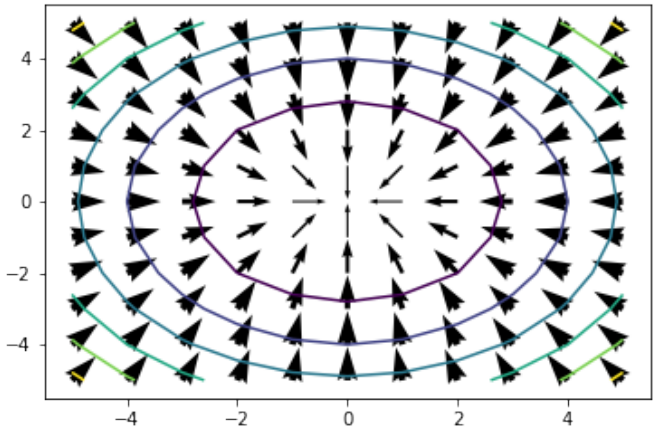
Plotting the Gradient

```
import matplotlib.pyplot as plt
import numpy as np
import math
## Create a uniform grid
el = np.arange(-5,6)
nx, ny = np.meshgrid(el, el, sparse=False, indexing='ij')
## flatten the gird to 1-d and compute the value of the function z
x_coord = []
y_coord = []
z = []
for i in range(11):
    for j in range(11):
        x_coord.append(float(-nx[i,j]))
        y_coord.append(float(-ny[i,j]))
        z.append(nx[i,j]**2 + ny[i,j]**2)
x_grad = [-2 * x for x in x_coord]
y_grad = [-2 * y for y in y_coord]
```

Plotting the Gradient

```
## Plot the arrows using  width for gradient
plt.xlim(-5.5,5.5)
plt.ylim(-5.5,5.5)
for x, y, xg, yg in zip(list(x_coord), list(y_coord), list(x_grad),
     list(y_grad)):
    if x != 0.0 or y != 0.0: ## Avoid the zero divide when scaling the arrow
        l = math.sqrt(xg**2 + yg**2)/2.0
        plt.quiver(x, y, xg, yg, width = l, units = 'dots')

## Plot the countours of the function surface
z = np.array(z).reshape(11,11)
plt.contour(el, el, z)
```

Plotting the Gradient



- The arrows in the plot point in the direction of the gradient.
- The width of the arrows is proportional to the value of the gradient. The width of the arrows and the gradient decreases as function gets closer to the minimum. If this is the case everywhere, you can say that a function is convex. It is always much easier to find minimum of convex functions.
- The direction of the gradient is always perpendicular to the contours.

Optimization

Optimization

- Goal of Machine Learning is Optimization
- Trying to find the“ best” model for prediction
- The “best” means that which minimizes prediction error.
- The technique, widely used is called as “Gradient Descent”

Example

Trying to find relation (formula/model) between input x and output y

x	y
2	4
6	12
3	7
4	7
12	24
21	40

Any guesses?

Example

How do you even start?
Assume $y = w.x$

x	y
2	4
6	12
3	7
4	7
12	24
21	40

Now, how to find w?

Example

Assume random value, say $w = 1$.
Using the relation, predict y, called y'

x	y	y'
2	4	2
6	12	6
3	7	3
4	7	4
12	24	12
21	40	21

Is y' matching actual y? NO. What to do? Need to find such w for which both match, right?

Example

How much OFF we are?

x	y	y'	y - y'
2	4	2	2
6	12	6	6
3	7	3	4
4	7	4	3
12	24	12	12
21	40	21	19

Too much of diff. Summation of squares (to nullify effect of sign) will give total error. Huge!!
Lets try another value for w and see if that works.

Example

Lets try $w = 0.5$.
Using the relation, and new w predict y find the error again

x	y	y'	y - y'
2	4		
6	12		
3	7		
4	7		
12	24		
21	40		

Error going up or down? UP too much. So out "direction" of change was not good. Instead of reducing w to 0.5, we should try increasing it to 1.5.

Example

Lets try $w = 1.5$.
Using the relation, and new w predict y find the error again

x	y	y'	y - y'
2	4		
6	12		
3	7		
4	7		
12	24		
21	40		

Error going up or down? Better. So out "direction" of change good but still error does not seem to close to 0. Increase w further.

Example

Lets try $w = 2$.
Using the relation, and new w predict y find the error again

x	y	y'	y - y'
2	4		
6	12		
3	7		
4	7		
12	24		
21	40		

Looks far better. Best? Probably. So relation now is $y = 2.x$. Seems to be correct for all rows except 1 or 2. But thats ok.

Maximization Minimization

- Ideas is to minimize the differences, also called as Error, Loss, Cost.
- It depends on w
- So, this is a minimzation problem of Cost wrt w
- How do you find minimum of ANY function?

Maximization Minimization

- Suppose we have a function f taking vector (or list of values) and outputs a single number.
- We frequently need to maximize or minimize the the function.
- Meaning, we wish to try many input vectors and find one where the result is minimum or maximum
- Examples: minimize cost of operations, maximize profits, etc.

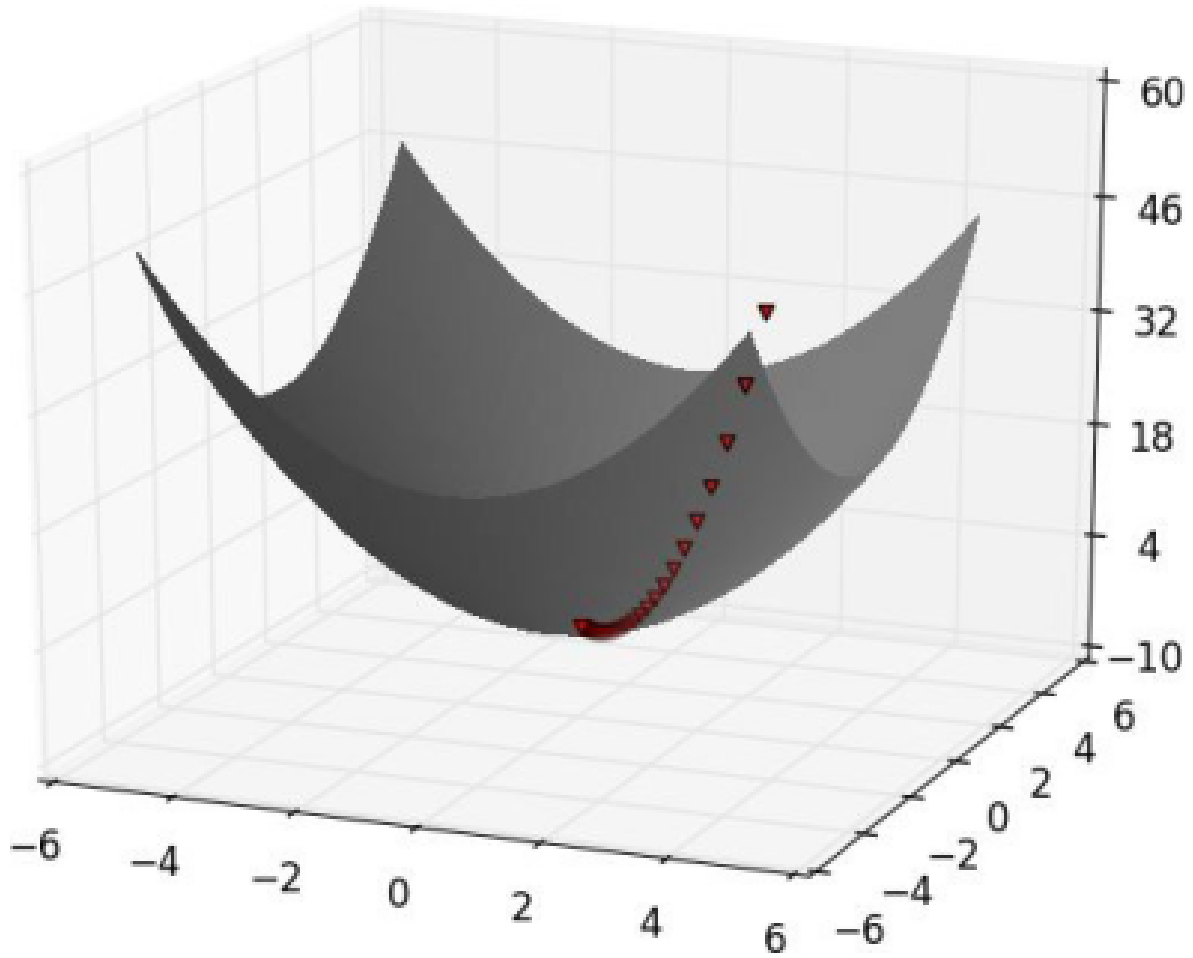
Example

Say, our function is $\sum x_i^2$

```
def sum_of_squares(v):
  """computes the sum of squared elements in v"""
  return  sum(v_i**2 for v_i in v)
```

- Let's use gradients to find the minimum among all three-dimensional vectors
- We'll just pick a random starting point
- Take tiny steps in the opposite direction of the gradient
- Until we reach a point where the gradient is very small.

Example



Example

Random vector, starting point

```
v = [random.randint(-10,10) for i in range(3)]
```

Our Gradient function is, derivative of $\sum x_i^2$ which is $\sum 2x_i$

```
def sum_of_squares_gradient(v):
  return  [2 * v_i for v_i in v]
```

Example

Step is computed as below, if size is -ve then its in opposite of gradient.

```
def step(v, direction, step_size):
  """move step_size in the direction from v"""
  return  [v_i - step_size * direction_i for v_i, direction_i in zip(v,
     direction)]
```

Example

Core logic:

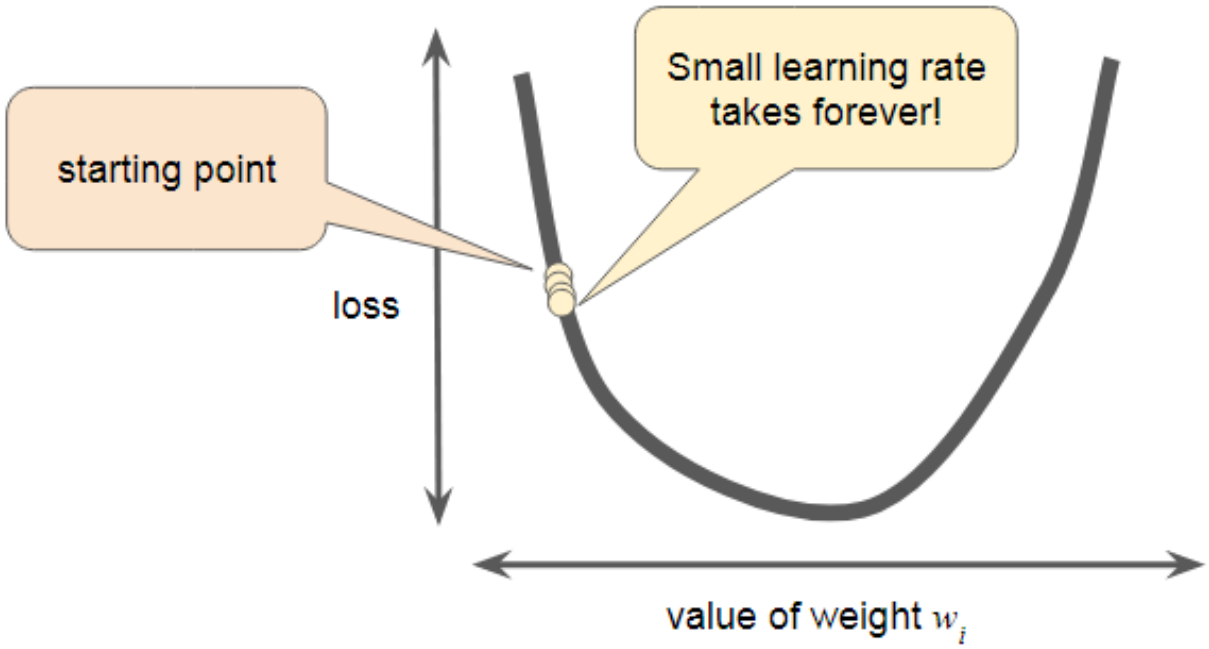
```
tolerance = 0.0000001
while True:
  gradient = sum_of_squares_gradient(v)
  next_v = step(v, gradient, 0.01)
  if gradient < tolerance:
    break
  v = next_v
```

We end-up at $[0, 0, 0]$ which is obvious.

Considerations

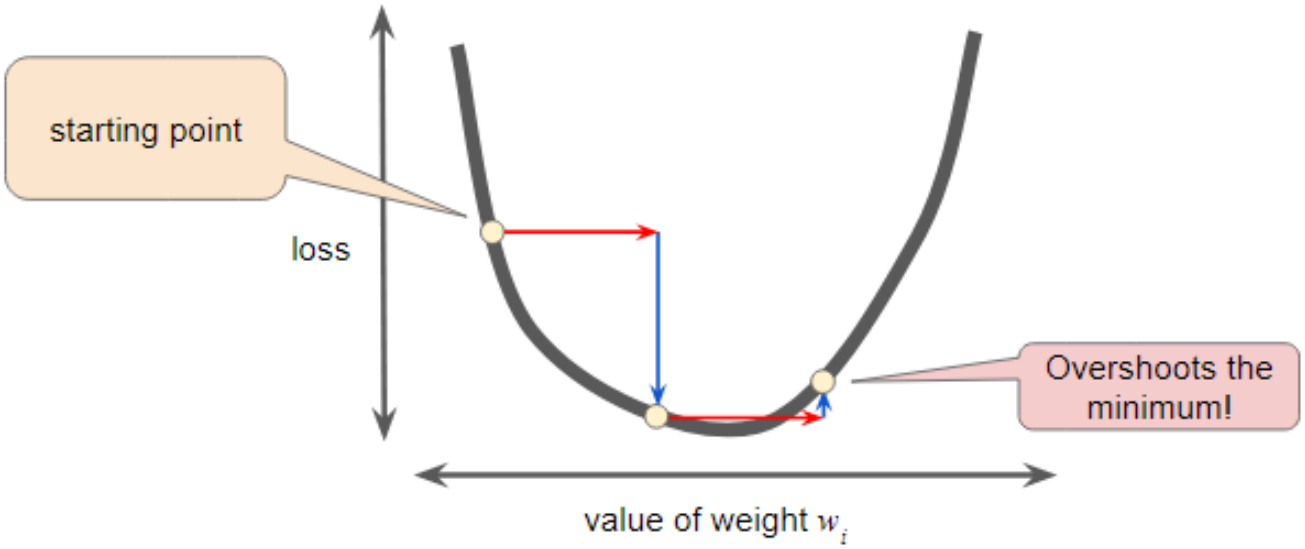
- Choosing the right step size is critical
- Using fixed step size
- Gradually shrinking step size

Learning rate is too small



(Reference:https://developers.google.com/machine-learning/crash-course/reducing-loss/learning-rate)

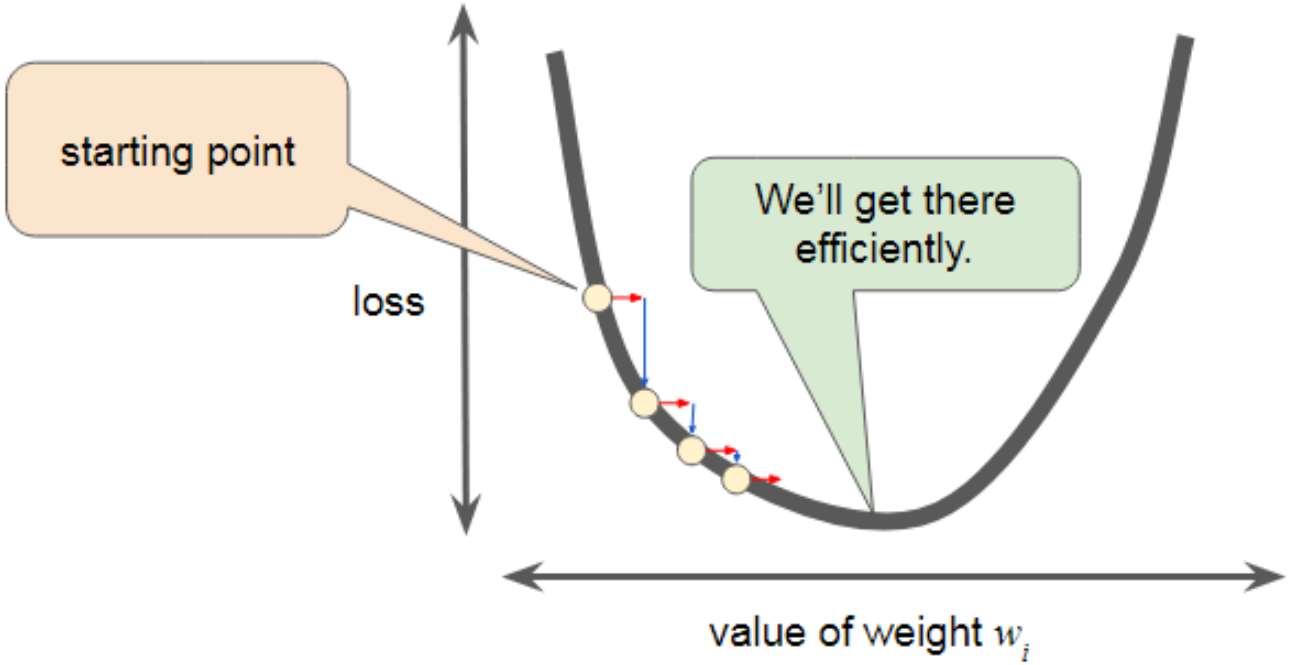
Learning rate is too large



(Reference:https://developers.google.com/machine-learning/crash-course/reducing-loss/learning-rate)

Learning rate is just right

The Goldilocks value is related to how flat the loss function is. If you know the gradient of the loss function is small then you can safely try a larger learning rate, which compensates for the small gradient and results in a larger step size.



(Reference:https://developers.google.com/machine-learning/crash-course/reducing-loss/learning-rate)

Calculus: Conclusion

We saw ...

- Calculus deals with infinitesimal processes.
- Limiting situations.
- Derivatives
- Gradients
- Optimization

Archimedes ...

- Archimedes had a calculus mindset
- He re-arranged discs, cylinders, cones, etc. to make "easy to measure" slices
- As you are doing the same here ...
- This will make him tear up!!

(Ref: Calculus, Better Explained: Summary - Better Explained)

Thanks ... yogeshkulkarni@yahoo.com