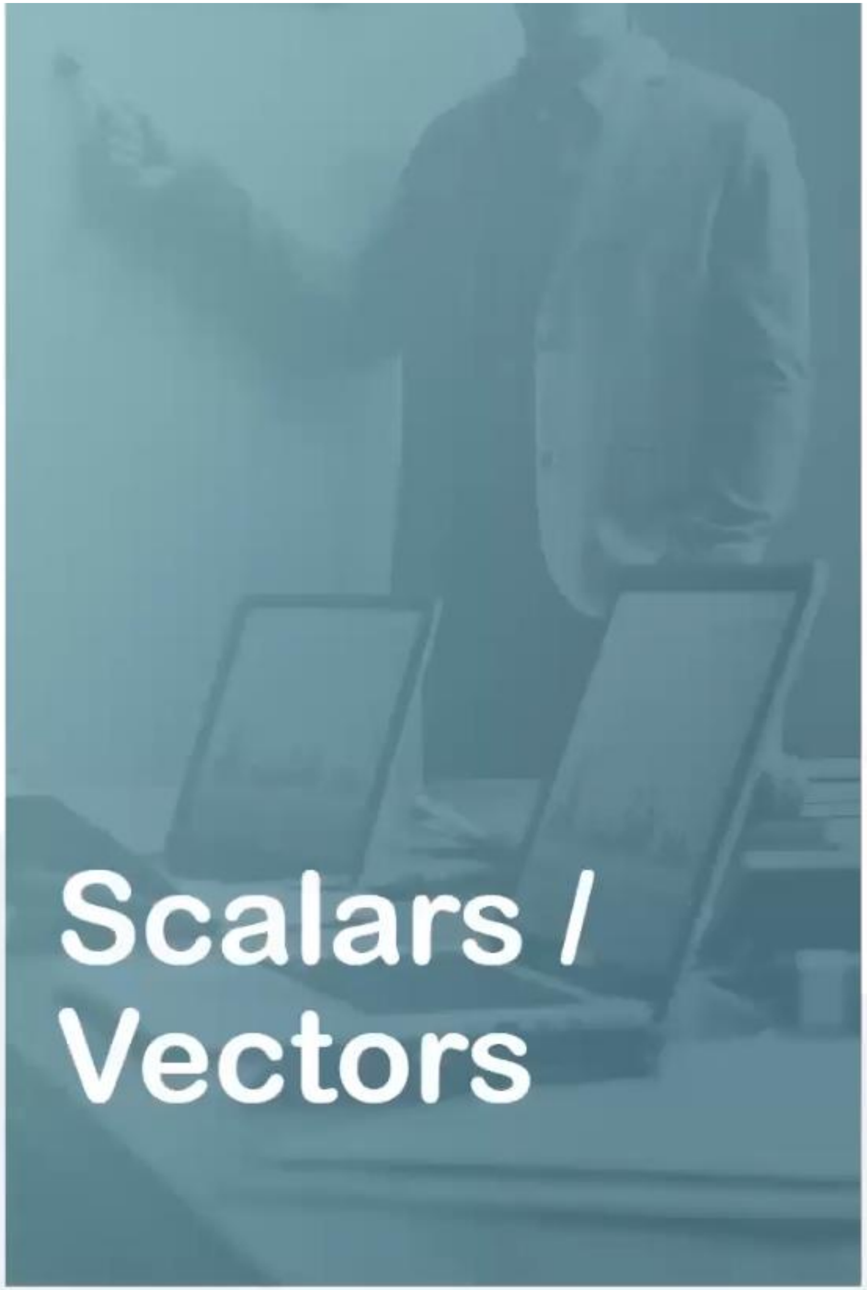
Mathematics for ML

Agenda..

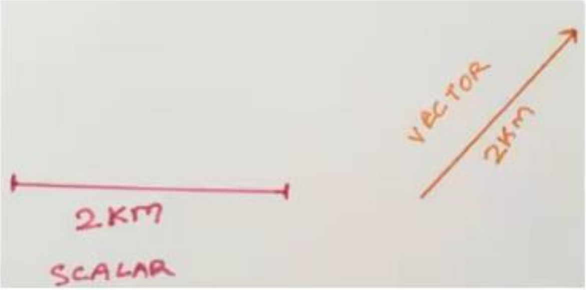
#	Topics
1	Scalars
2	Vectors
3	Matrices
4	Calculus



Scalars / Vectors

Why is Data represented as a 'Vector' in Data Science Problems?

- Science and Maths are used to describe the world around us.
- There are many quantities which require only 1 measurement to describe them. e.g Length of a string, or area of any shape or temperature of any surface. Such quantities are called scalars.
- Any quantity which can be represented as a number (positive or negative) is called scalar. This value is known as magnitude.
- On the other hand, there are quantities which require at least 2 measurements to describe them. Along with the magnitude, they have a "direction" associated e.g velocity or force.
- These quantities are known as "Vectors".
- When we say that a person ran for 2 Kms, its a scalar but when we say that a person ran for 2 Kms, North-east from his initial position, its a vector.

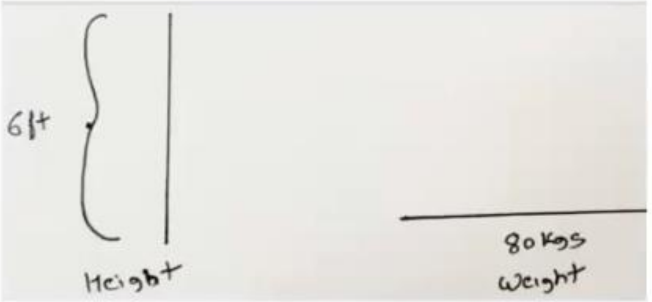


Scalar Quantity

Let's say you are collecting some data about a group of students in a class. You are measuring the height and weight of each student and the data collected for 5 students is as follows:

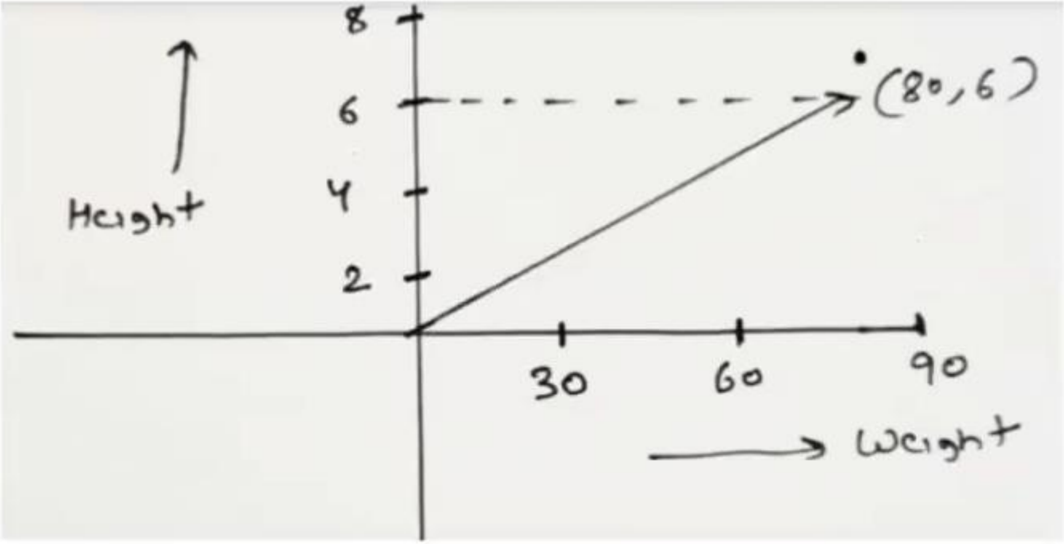
Heigt (in feets)	Weight (in Kgs)
6	80
5.4	54
5	50
5.7	65
5.8	72

Each individual measurement here is a scalar quantity. So, height or weight viewed stand-alone are scalars.



Vectors

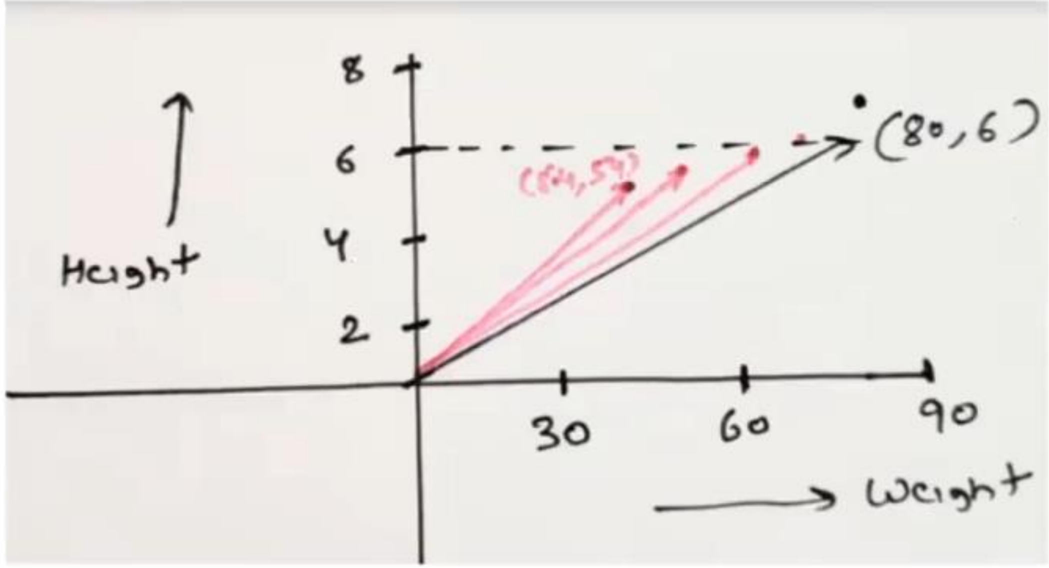
However, when you look at the observation about each student i.e height **and** weight together for every student, you can think of it as a vector



- In this representation, each student is a vector. It is described by 2 characteristics: height and weight.
- Now, there is no real "direction" here in this vector as per the standard definition.
- But when we represent quantities in more than one dimension (in this case, it's 2 — height and weight), there is a sense of orientation with respect to each other.

Representing Multiple Vectors

So, when we are looking at 5 students (5 vectors), they have a magnitude and an idea of "direction" with respect to each other.



Since most useful datasets always have more than 1 attribute, they all can be represented as "Vectors". Every observation in each data set can be thought of as a vector. All possible observations of a data set constitute a "vector space".

The benefit of representing data as vectors is — we can leverage vector algebra to find patterns or relationships within our data

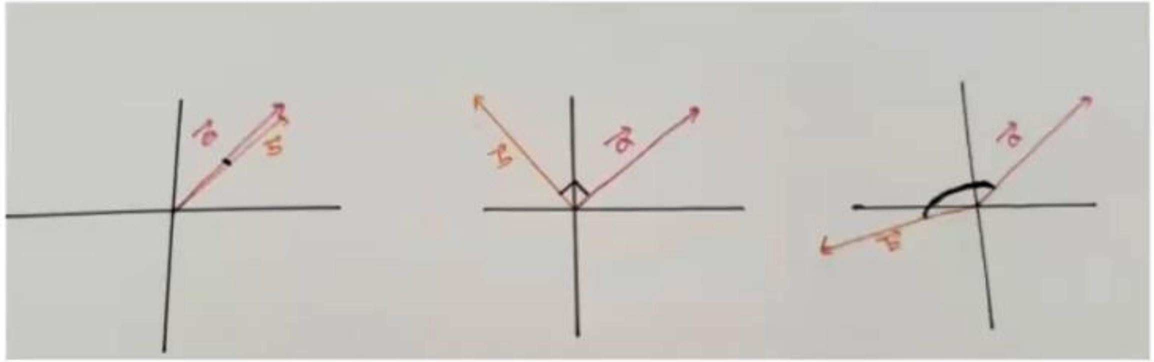
How can vectors help?

- Looking at the vectors in vector space, you can quickly compare them to check if there is a relationship. For example, you deduce that similar vectors will have smaller angle between them i.e their orientation will be close to each other.

- In our sample data, the students (5.4,54) and (5, 50) are quite similar.

- The angle between the vectors indicates “similarity” between them.

- The vectors in the same direction (close to 0 degrees angle) are similar while vectors in the opposite direction (close to 180 degrees angle) are dissimilar. Theoretically. if the vectors are at 90 degrees to each other (orthogonal), then there is no relationship between them.



Three Dimensional Vector

Let's take the example of v, a three-dimensional vectors defined as follows:

Equation

$$\mathbf{v} = \begin{bmatrix} \mathbf{3} \\ \mathbf{4} \\ \mathbf{2} \end{bmatrix}$$

As shown in in Figure 5, you can reach the endpoint of the vector by traveling 3 units in the xx-axis, 4 in the yy-axis, and 2 in the zz-axis.

More generally, in a n-dimensional space, the position of a terminal point is described by n components.

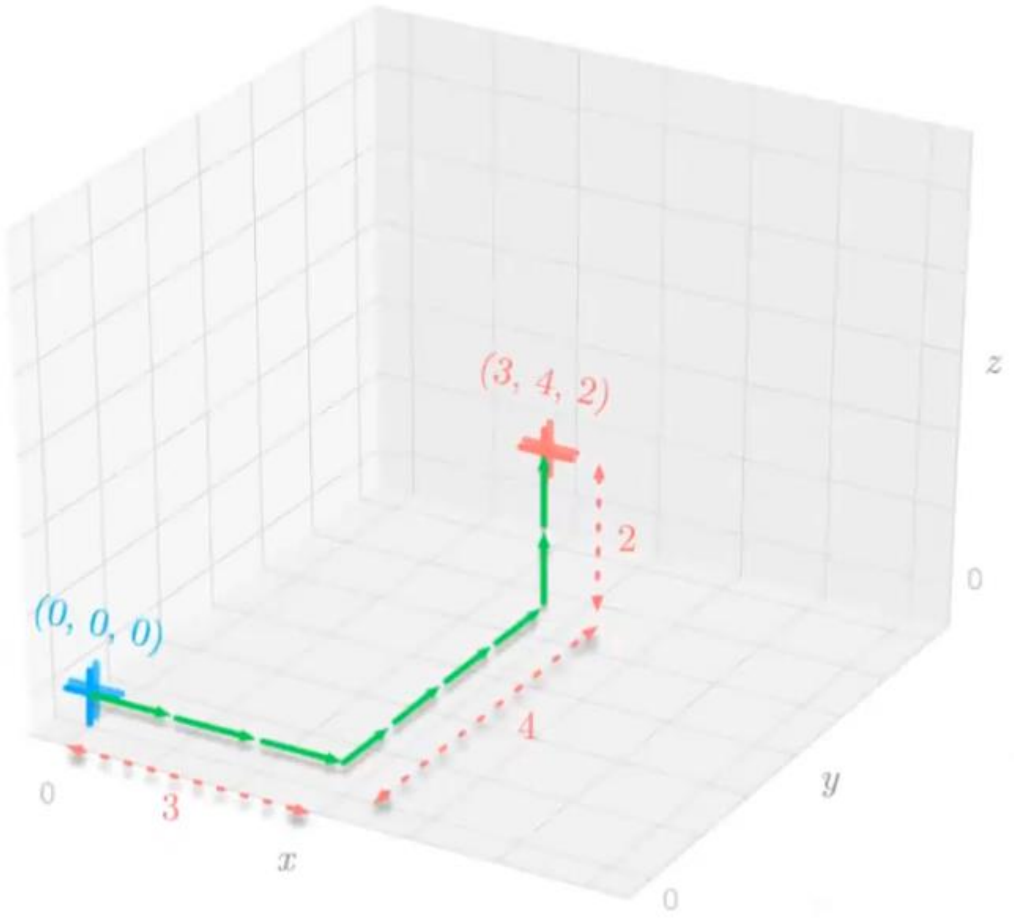


Figure 5: Three-dimensional representation of the origin at (0, 0, 0) and the point at (3, 4, 2).

Is there a mathematical function which can capture this relationship?

Think of the trigonometric function — cosine.

Cosine Similarity

Cosine Similarity is a metric that gives the cosine of the angle between vectors. It can be calculated using the "dot product" of 2 vectors. Mathematically,

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos\theta$$

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

Degrees	0^o	90^o	180^o
Cosine	1	0	-1

So now whenever you hear about vectors in Data Science, think about the dot product and cosine of the angle between them.

You will be looking at how similar or dissimilar those vectors are.

Text Vectorization

The process of converting or transforming a data set into a set of Vectors is called Vectorization. It’s easier to represent data set as vectors where attributes are already numeric. What about textual data?

“Word Embedding” is the process of representing words or text as Vectors

There are a number of techniques of converting/representing text as Vectors. One of the simplest methods is Count Vectorizer. Below is a snippet of the first few lines of text from the book “A Tale of Two Cities” by Charles Dickens:

It was the best of times,
it was the worst of times,
it was the age of wisdom,
it was the age of foolishness

How can the above 4 sentences be converted into vectors

Step 1: Get unique words from the collection of your text. The total text you have is called “corpus”

The unique words here (ignoring case and punctuation) are:

```
['age',
 'best',
 'foolishness',
 'it',
 'of',
 'the',
 'times',
 'was',
 'wisdom',
 'worst']
```

Some Important Rules

This is a vocabulary of 10 words from a corpus containing 24 words.

Step 2: For each sentence, create a list of 10 zeroes

Step 3: For each sentence, start reading the word one by one. For each word, count total occurrence in the sentence. Now identify the position of the word in vocabulary list above and replace the zero with this count at that position.

For our corpus, the vectors we got are:

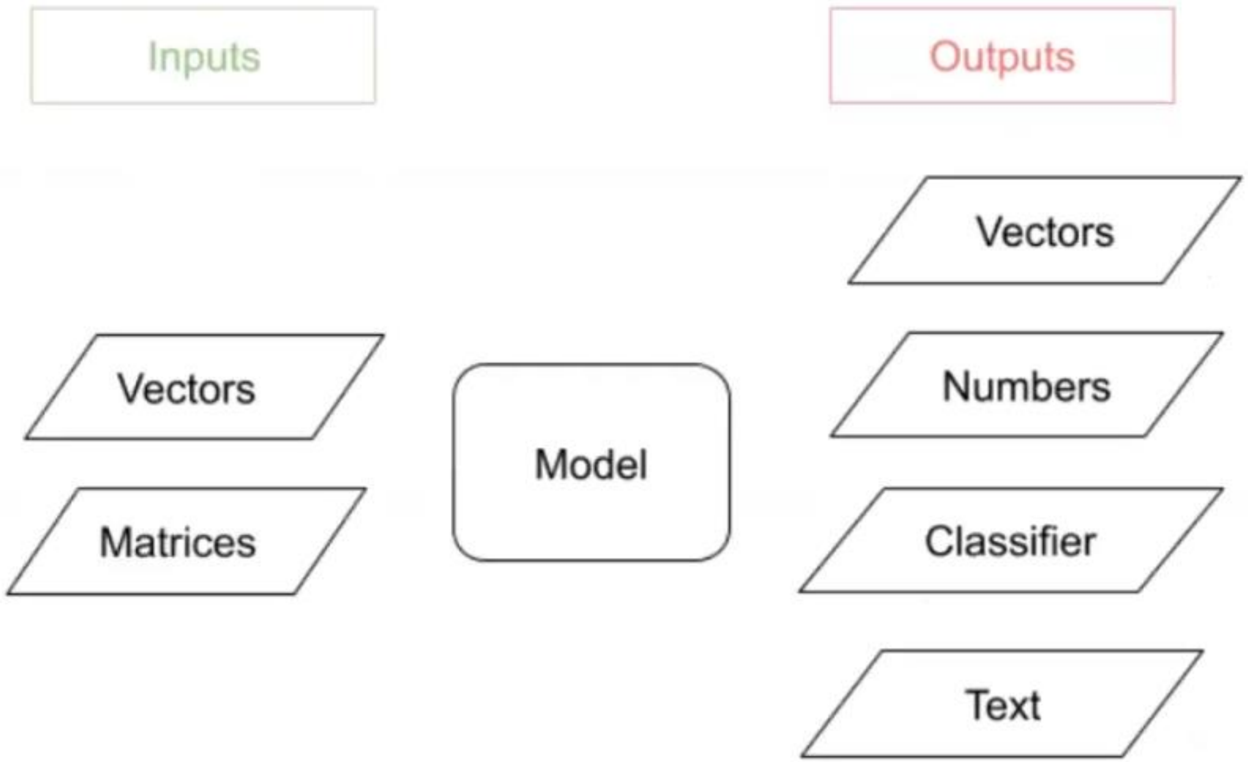
“It was the best of times“= [0 1 0 1 1 1 1 1 0 0]

“it was the worst of times” = [0 0 0 1 1 1 1 1 0 1]

“it was the age of wisdom” = [1 0 0 1 1 1 0 1 1 0]

“it was the age of foolishness” = [1 0 1 1 1 1 0 1 0 0]

Use of Vectors



Use of Vectors

There are three different stages. Each one has a slightly different use case for vectors. In this article, we'll clear this all up by looking at vectors in relation to these stages:

Input: Machines can't read text or look at images like you and me. They need input to be transformed or encoded into numbers. Vectors, and matrices (we'll get to these in a minute) represent inputs like text and images as numbers, so that we can train and deploy our models. We don't need a deep mathematical understanding of vectors to use them to encode information for our inputs. We just need to know how vectors relate to features, and how we can represent those features as a vector.

Model: The goal of most ML projects is to create a model that performs some function. It could be classifying text, or predicting house prices, or identifying sentiment. In deep learning models, this is achieved via a neural network where the neural network layers use linear algebra (like matrix and vector multiplication) to tune your parameters. This is where the mathematical definition of vectors is relevant for ML.

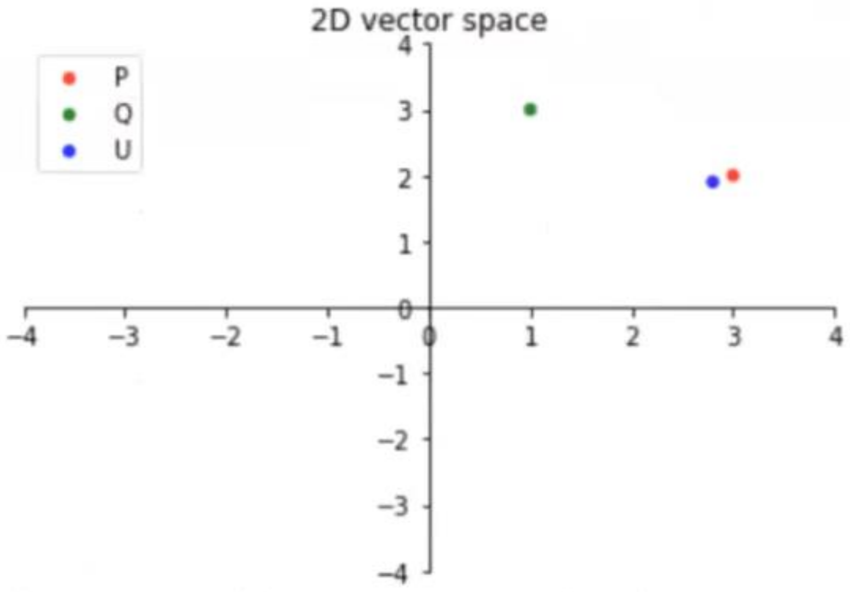
Use of Vectors

Output: The output of our ML model can be a range of different entities depending on our goal. If we're predicting house prices, the output will be a number. If we're classifying images, the output will be a category of image. The output, however, can be a vector as well. For example, NLP models like the Universal Sentence Encoder (USE) accept text and then output a vector (called an embedding) representing the sentence.

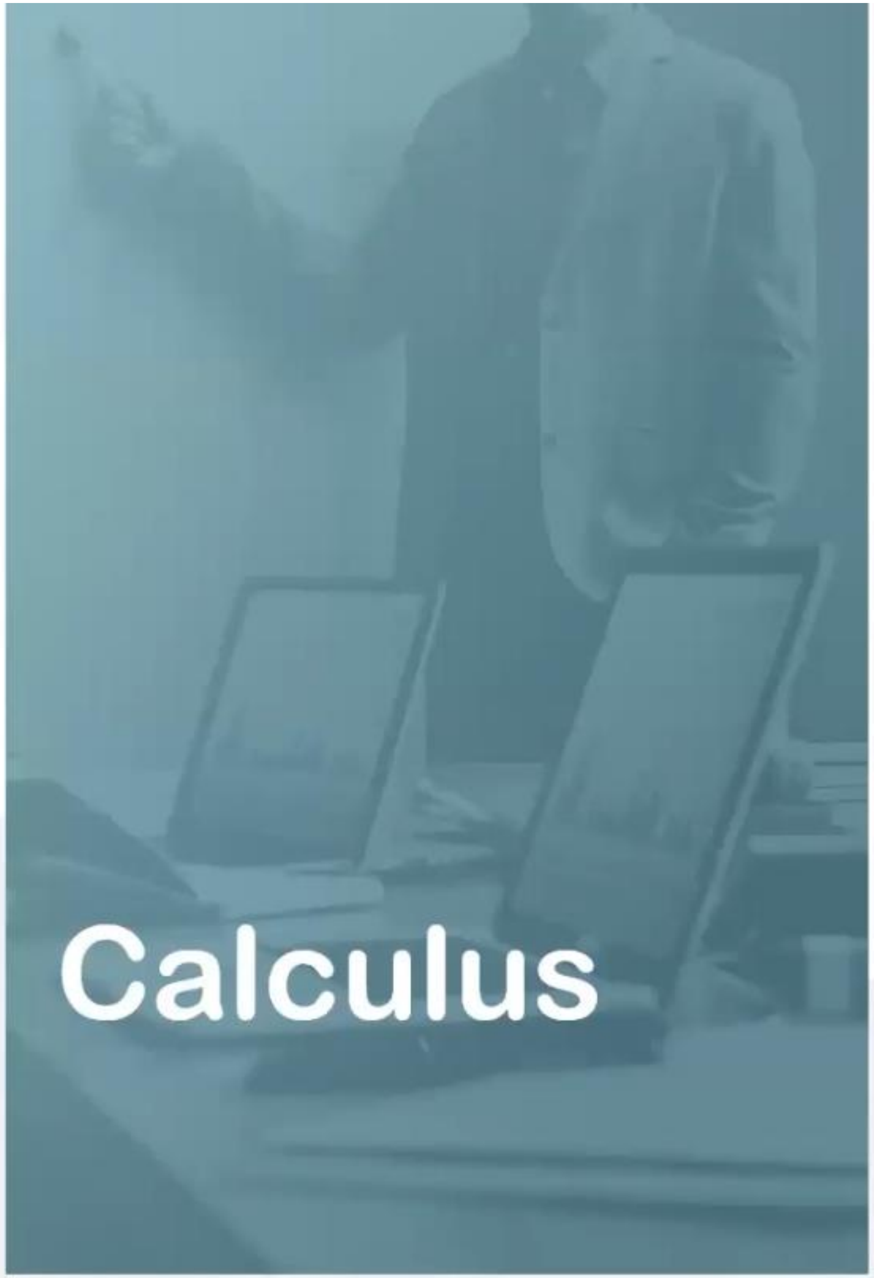
You can then use this vector to perform a range of operations, or as an input into another model. Among the operations you can perform are clustering similar sentences together in a vector space, or finding similarity between different sentences using operations like cosine similarity.

Vector Spaces

We can think of a vector space as the area which contains all of our vectors. We can think of these spaces in relation to the dimensions of the vector. If our above example of a vector had two dimensions, it could be described by X and Y coordinates like most 2-D graphs you would draw on matplotlib:



2-D vector spaces are similar to common graphs we would draw in matplotlib. Notice here that we have drawn the vectors as points without the arrows. This is to show that we can also think of vectors as points in these vector spaces.


Calculus

Calculus – Mathematics of Change

- Calculus is the branch of mathematics that deals with study of change
- Calculus helps in finding out the relationship between two variables (quantities) by measuring how one variable changes when there is a change in another variable and how these changes accumulate over time.

Calculus – Mathematics of Change

In ML, The understanding or learning is often expressed as a mathematical function that captures the relationship between the entities or dimensions involved. e.g let's say you are driving a car. You are measuring the speed of the car as time is passing by. Some sample measurements can be as follows:

As seen, there is a relationship between speed and time. As time passes by, the speed of the car increases at a constant rate. Every minute, the speed of the car increases by 5 times. This relationship can be expressed as follows:

Time (in minutes)	Speed(in Km/hr)
0	0
1	5
2	10
3	15
4	20
5	25

Speed is a Function of time

Speed = f(time)

S = 5* t

Constant Change (Linear)

Visually, you can see that there is a linear relationship between speed and time

You can imagine that this relationship is representing the equation of a line:

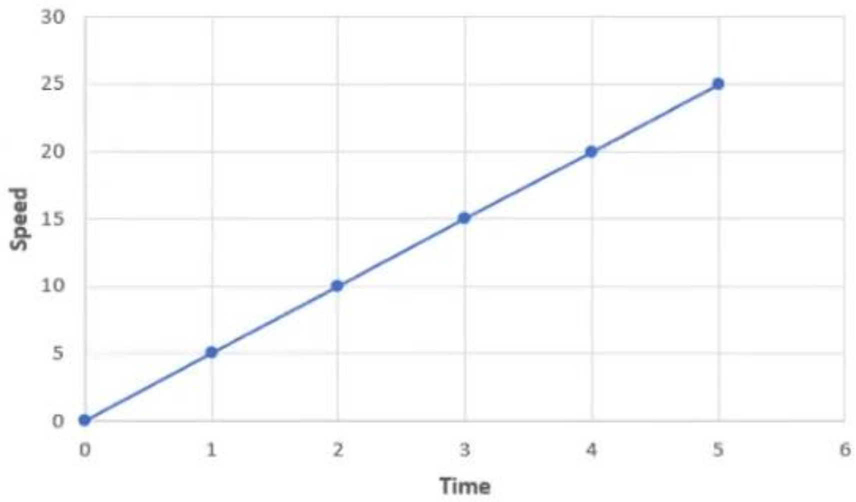
$y = mx + b$

where m is the slope and b is the intercept of the line.

In our example, the equation becomes $s = 5t + 0$

Focus on this thing known as "Slope". What does it represent? We know that as time goes by, the speed of the car is changing. By what factor is this changing?

We can find out by checking how much speed has changed for a change in time. e.g between t = 1 and t =2, the speed has changed from 5 to 10.



Constant Change - Linear

This can be written as:

$$\Delta s = \text{Change in Speed} = 10 - 5 = 5$$

$$\Delta t = \text{Change in Time} = 2 - 1 = 1$$

When things change, a more useful measurement to make is “Rate of Change”. The rate of change gives us an idea of how much one quantity is dependent on the other. If the rate is more, there will be a large correlation between the quantities and vice-versa.

In our case, if we divide the “change in speed” with “change in time”, it gives us the rate of change

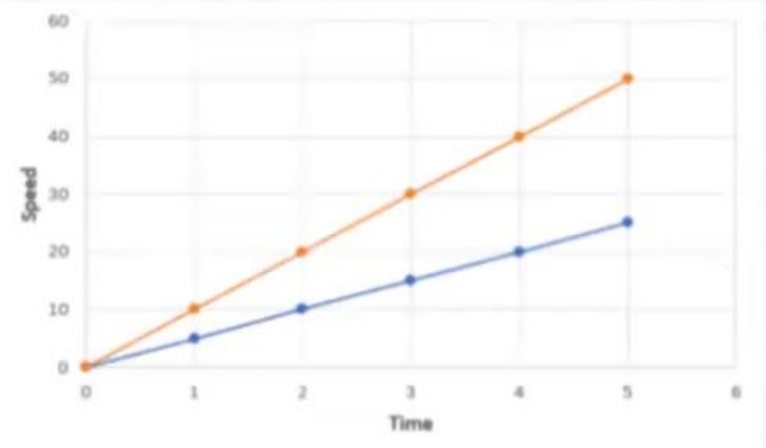
$$\frac{\Delta s}{\Delta t} = 5$$

This is the rate of change of speed with respect to change in time. For our example, this rate is constant. You can take other time intervals (e.g t=2 and t=5) and measure the change in speed (25–10). The rate is the same — 5 ((25-10)/(5–2)) = 15/3=5

Slope / Gradient

- The rate of change is also known as slope or gradient.
- Notice that greater the rate of change, greater will be the inclination of the line (hence, greater the slope).
- Compare the slope of the orange and blue lines in the below graph:

Time (in minutes)	Speed1	Speed 2
0	0	0
1	5	10
2	10	20
3	15	30
4	20	40
5	25	50

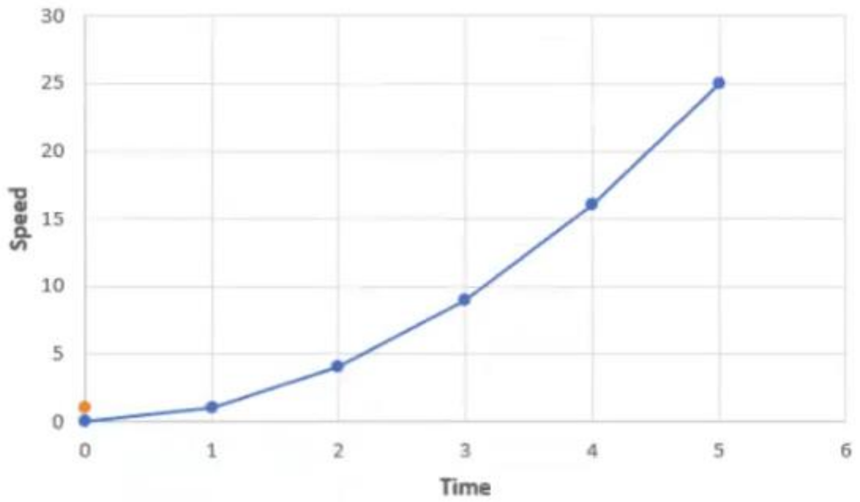


Non-Constant Change (Non-Linear)

In our example above, the speed of the car is increasing at a constant rate. What will happen if this rate of change is not constant? What if the car is accelerating and the rate of change in speed is different every minute

How to measure the rate of change, in this case, because its different for different time periods? We can not define a single rate of change in speed for this kind of data.

Time (in minutes)	Speed
0	0
1	1
2	4
3	9
4	16
5	25
6	36



You can notice that the rate of change of speed climbs up as time passes. The mathematical function that captures this relationship is not a straight line but its a “curved line”.

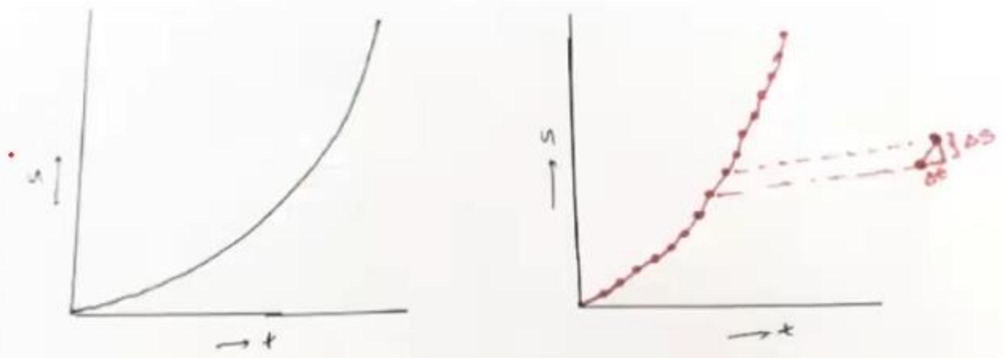
$$s = t^2$$

Slope / Gradient

From our earlier example, we know that the slope or gradient of a straight line gives us the rate of change of the Y-axis variable (Speed) with respect to the X-axis variable(Time). The property of “straightness” means that the rate of change was constant.

But…

How to measure the “slope” or “inclination” of a line which is curved?
The property of “curviness” means that rate of change is not constant.



Here comes the beauty of calculus…….

What you can do is to imagine the curve as a collection of lots of “very small straight line segments”

Slope / Gradient

We will now be calculating the slope of a “very small” line segment of the curve. This slope represents a “very small” change in speed with respect to a “very small” change in time.

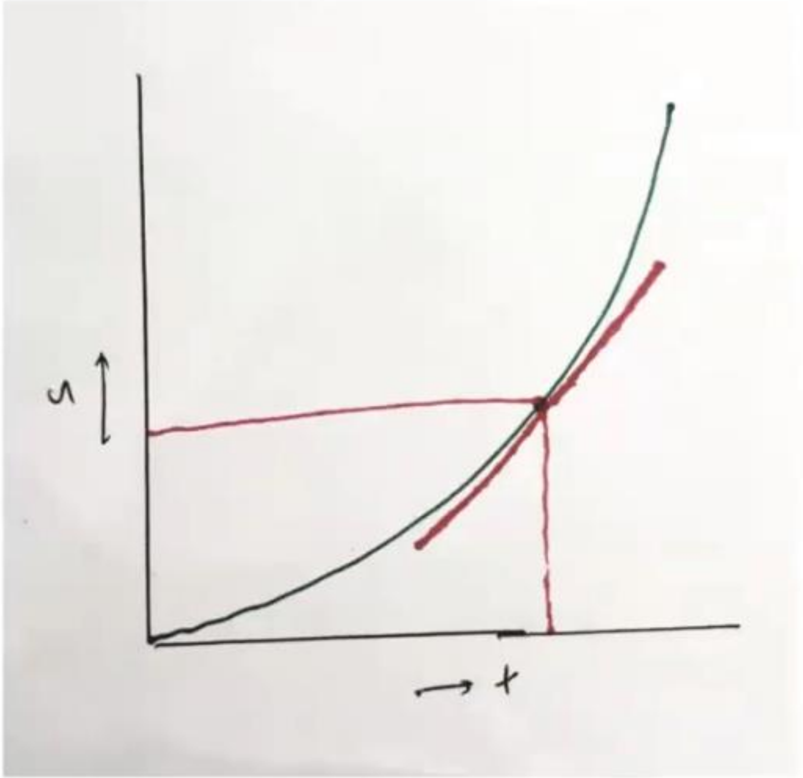
Let's denote “very small” change in speed as “ds”.

Similarly, a “very small” change in time as “dt”

The slope of this “very small” line segment is $\frac{ds}{dt}$

Slope / Gradient = Tangent

Theoretically, we can make ds and dt infinitesimally small, so that this “very small” line segment” becomes a point. The rate of change or slope at this point or instant of time is a tangent to the curve.



So, the function and rate of change of this function at a particular instant, in our example, is

$$s = t^2$$

$$\frac{ds}{dt} = 2t$$

Slope of a line parallel to X Axis

In calculus terminology, the process of finding out the rate of change of a variable with respect to another is known as “differentiation”. In other words, if

y = f(x)

then the process of finding out dy/dx is known as “differentiation” or “differentiating”. The ratio dy/dx is called “the differential coefficient of y with respect to x” or “derivative”. Remember, its nothing but the rate of change of y with respect to x.

Keep in mind — “dx” means a very small part of x

What if ‘y’ is a constant. What will be dy/dx in this case? It will be 0 as the value of y is not changing. It's a constant.

$$\frac{d\,(constant)}{dx} = 0$$

Application of Differential Calculus in Machine Learning

The most robust application of Calculus in Machine Learning is the Gradient Descent algorithm in Linear regression (and Neural Networks)

Linear regression involves using data to calculate a line that best fits that data, and then using that line to predict scores on one variable from another. Prediction is simply the process of estimating scores of the outcome (or dependent) variable based on the scores of the predictor (or independent) variable. To generate the regression line, we look for a line of best fit. A line that can explain the relationship between the independent and dependent variable(s), better is said to be the best-fit line. The difference between the observed value and the actual value gives the error. The formula to calculate this error is also called the cost function.

The line of best fit will be expressed mathematically as

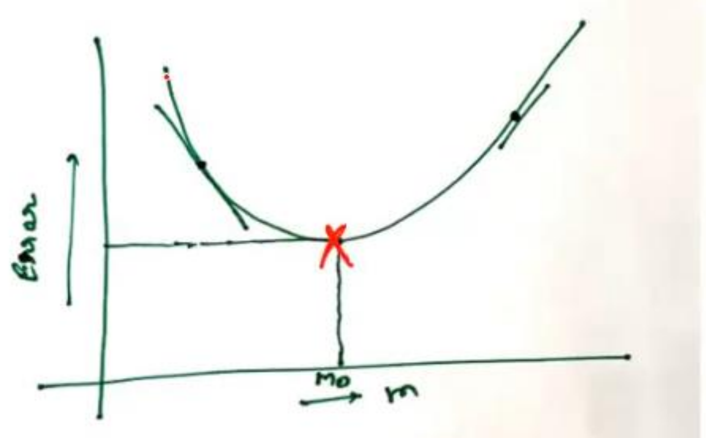
$Y = m.x + c$

where m and c are the slope and intercept of the line respectively. These are the two coefficients that the gradient descent algorithm has to find out.

Some Important Rules

The error will depend upon the coefficients of the line. If the value of coefficients is not optimal, then the error will be more. The cost function or the amount of error of Linear regression model is dependent upon the value of coefficient chosen.

This is where the calculus is used. We can find the rate of change of error with respect to the different values of coefficients. The value for which the rate of change is minimum(i.e 0, the bottom of the curve) is the optimal value.



So differentiation is used to find the minimum point of the curve. This point gives the optimal value of the coefficient that was being looked for!

Stochastic Gradient Descent

Finding Minima of a Loss Function…

Loss function is minimized to reduce error [Actual, Predicted]

Stochastic means Random

This is done by Stochastic Gradient Descent Method

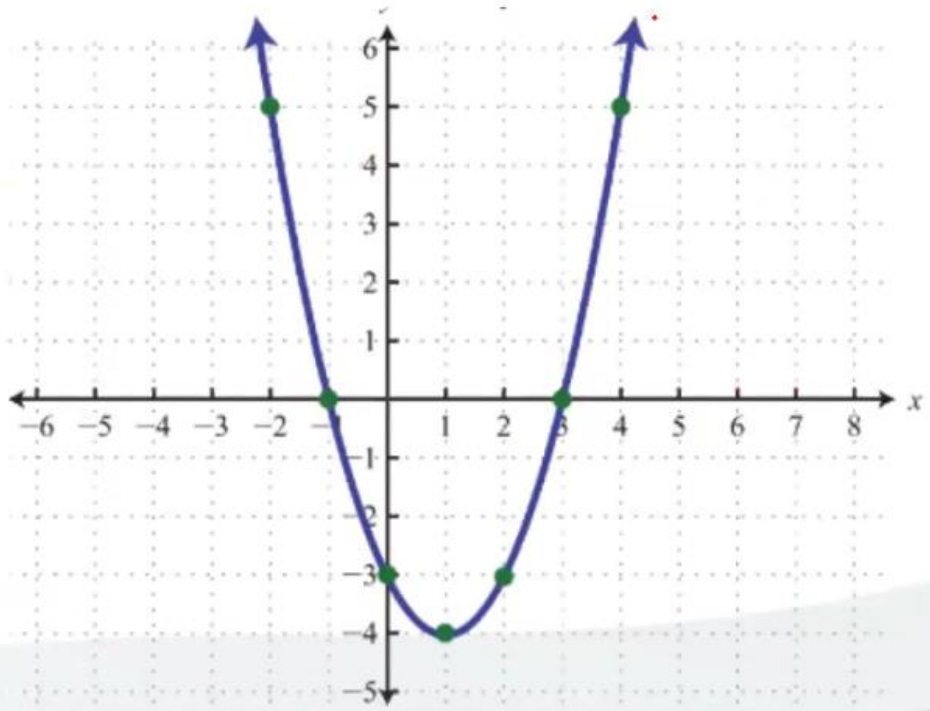
Objective is to find minima of a Loss/ Objective Function

- **Gradient:** Gradient means slope
- Gradient descent means descending a slope to reach lowest point on that surface.

Stochastic Gradient Descent Method…

- **Consider an Objective Function**
- Gradient Descent is an iterative algorithm, that starts from a random point on a function and travels down its slope in steps until it reaches the lowest point of that function.
- The objective of gradient descent algorithm is to find the value of "x" such that "y" is minimum.

Objective Function → $y = x^2 - 2x - 3$

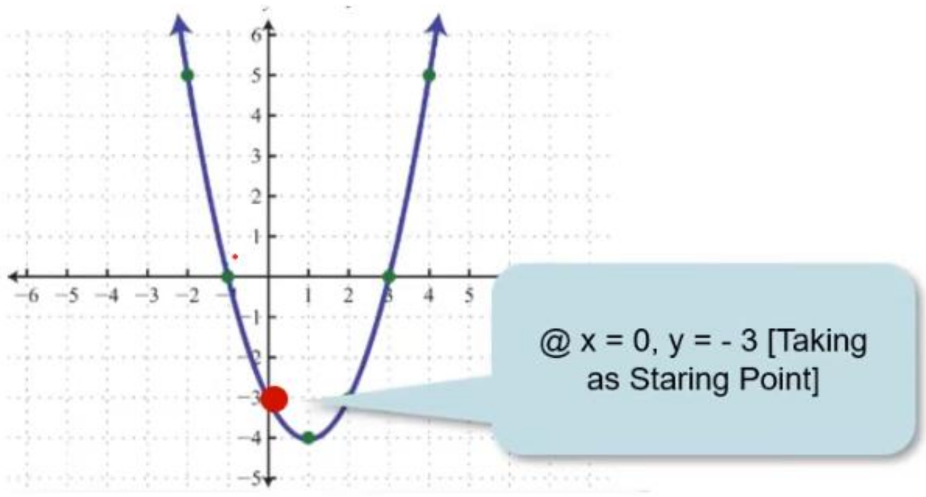


Steps in Stochastic Gradient Descent Method (1/4)…

- **Step 1:** Find the slope (differentiate "y" with respect to "x") of the objective function with respect to 'x'.

- **Differential of** ($y = x^2 - 2x - 3$)
$\Rightarrow$ **2x - 2 (Gradient)**

- **Step 2:** Randomly Initialize the Objective function (Assigning Say x=0). This gives us a point (x,y) → (0, -3) on the parabola. **[This becomes our starting point]**



Steps in Stochastic Gradient Descent Method (2/4)…

- **Step 3:** Putting value of starting points in Gradient ie x = 0

$\Rightarrow$ **2x - 2 (Gradient)**

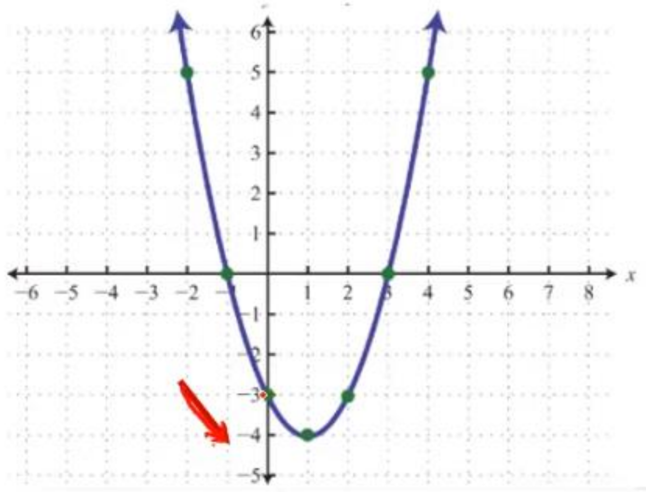
For x = 0; Gradient = -2

Gradient is –ve. Hence we have to move downwards

- **Step 4:** Find **Step Size**

Step Size = (Gradient) x (Learning Rate)

Learning Rate is a tuning value. It is usually taken as a very low value



Steps in Stochastic Gradient Descent Method (3/4)…

- **Step 4 (Continued):**

Step Size = (Gradient) x (Learning Rate)

$\Rightarrow$ **Step Size = (-2) x (0.01)**
$\Rightarrow$ **Step Size = -0.02**

Taking Learning Rate to be 0.01

- **Step 5:** Finding **New value of x**

New value (of x) = Old value (of x) – (Step Size)

$\Rightarrow$ **New Value (of x) = (0) - (- 0.02)**
$\Rightarrow$ **New Value (of x) = 0.02**

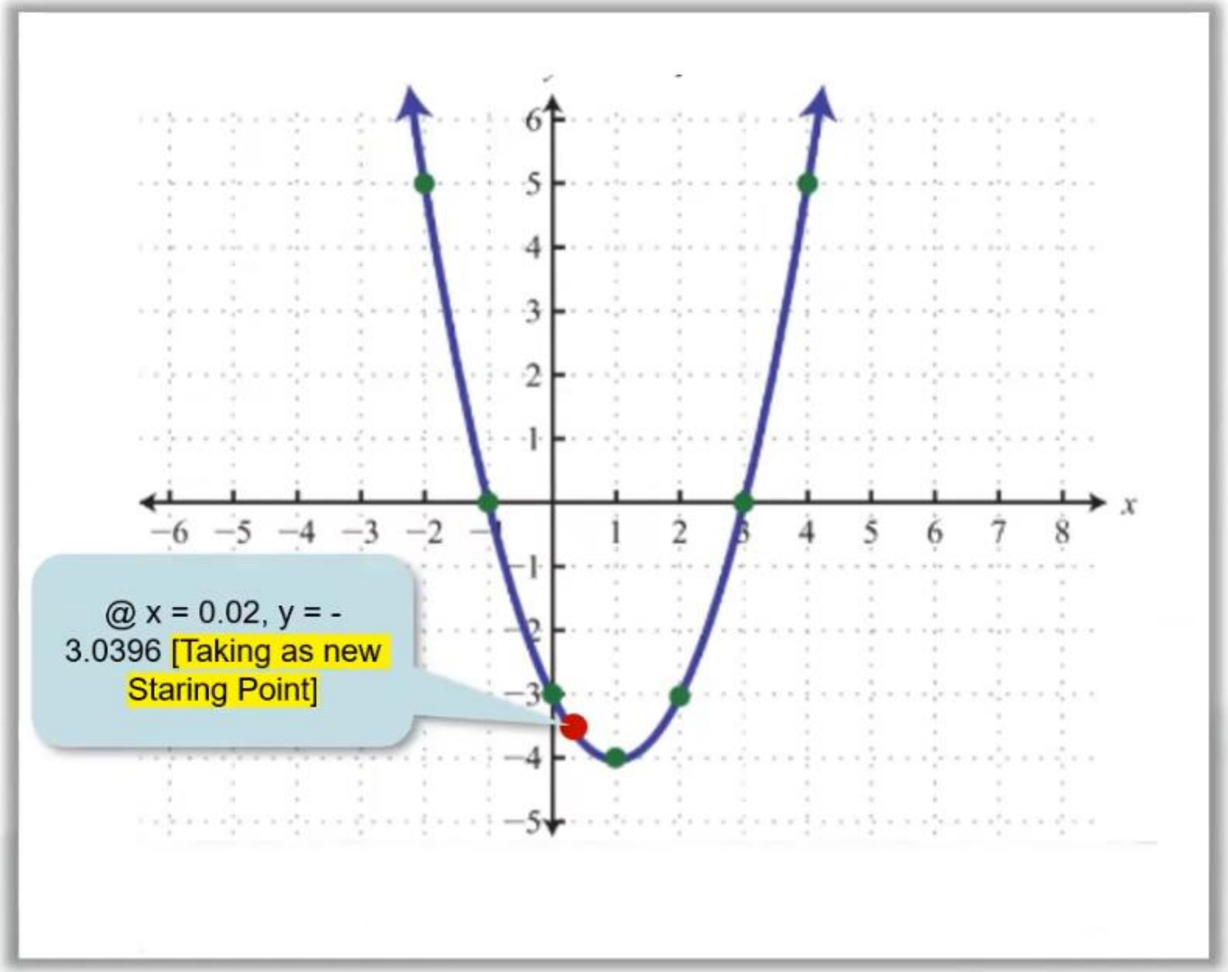
Steps in Stochastic Gradient Descent Method (4/4)…

- **<u>Step 5 (Continued):</u>** Finding new value of y
- Now y = 3.0396 [Taking as new Staring Point]

Now repeat Steps 3, 4, 5 until we reach the minima where slope/ gradient is 0

In our example, that point can be easily identified. For Gradient 2x-2 → @x=1, we will get the slope as **zero**

In reality, the Loss functions are complex and involves matrix calculation son tensors



Mini Batch Stochastic Gradient Descent is a slight variant of the SGD. Here instead of taking a single sample at a time, a batch of samples are taken, and the parameters are updated.

Integration

There is a second part to Calculus — Integration. It is simply, the reverse of Differentiation.

In differentiation, we break up things into smaller and smaller parts.

In integration, we accumulate or add up, all the smaller parts together. The symbol for integration is:

$$\int$$

Whenever you see this symbol, just replace it in your mind by “Add Up all”.

$\int dx$ = Add up all “small parts of x”

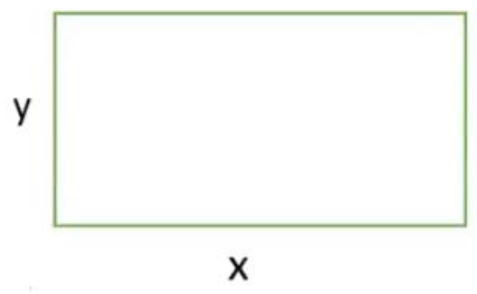
What will be the sum of all small parts of x?

$$\int dx = x$$

Why integrate? The simple answer

There are many things that can not be understood until and unless you break them up into smaller parts, do some operation for each smaller part and then accumulate or add up the results.

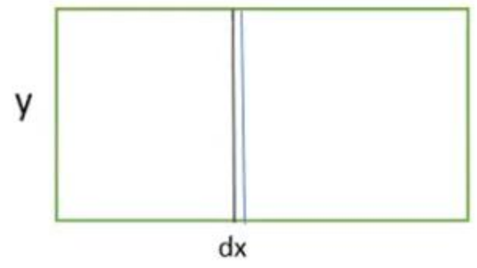
Let us take a rectangle:



We know that area of a rectangle is Length * Width (y *x)

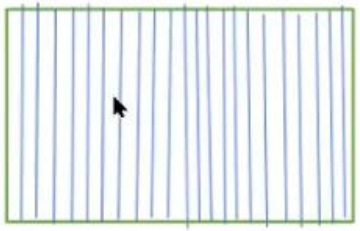
Can you prove it?

Think in terms of calculus. Imagine a smaller rectangle by taking a small bit of width(dx).



Some Important Rules

The whole of rectangle can be thought of sum of all smaller rectangles where width is dx



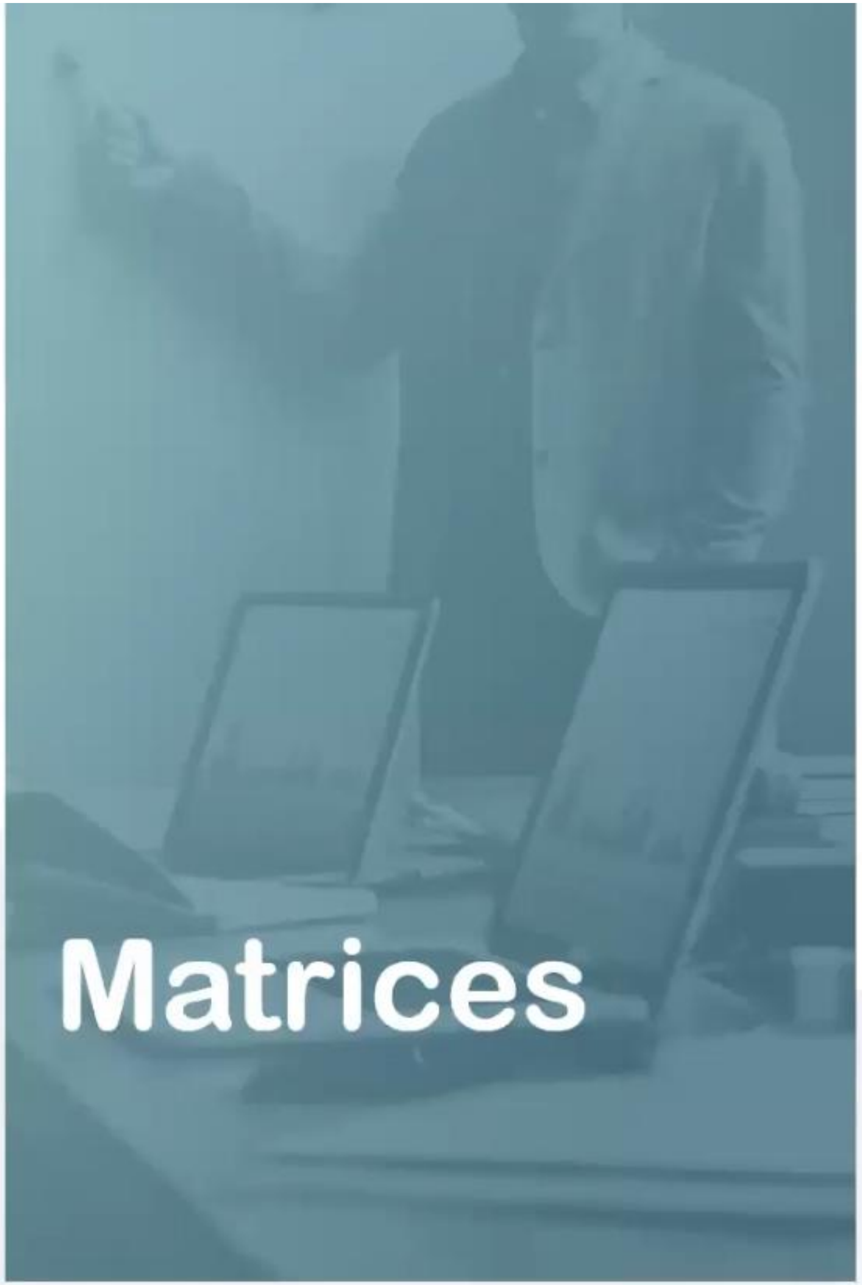
So, now we have ‘x’ number of small rectangles, each with width dx and length as y.

If you imagine dx to be very very small, the mini rectangle will eventually be reduced to a line (width close to 0 and length as y). The whole rectangle is just a collection of ‘lines’ each of length ‘y’. How many lines? — x.

Total area = Length of Line1 + Length of Line 2 + …….+ Length of Line X

$$Area = y \int (dx)$$

$$\text{Area} = Y * X$$



Matrices

Scalar, Vector & Matrix

Scalar	:	one number
Vector	:	row/column of numbers
Matrix	:	many rows and columns

Vectors

A vector is a 1 Dimension array of values

We use the notation $A \in \Re^n$ to denote that x is an n-dimensional vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

We use the notation x_i to denote the i^{th} entry of x

By default, we consider vectors to represent *column vectors*.

Matrices

A matrix is a 2 Dimension array of values

Notation: $A \in \Re^{m \times n}$ (with m rows and n columns)

$$\begin{array}{cl} A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}_{m \times n} & \begin{array}{l} \leftarrow \text{(Row 1)} \\ \\ \leftarrow \text{(Row m)} \end{array} \\ \begin{array}{cc} \uparrow & \uparrow \\ \text{(Column1)} & \text{(Column n)} \end{array} & \end{array}$$

Matrices

Column Matrix

$$\begin{bmatrix} 2 \\ 3 \\ 1 \\ 5 \end{bmatrix}_{4x1}$$

Only 1 column

Row Matrix

$$\begin{bmatrix} 6 & 1 & 4 \end{bmatrix}_{1x3}$$

Only 1 row

Square Matrix

$$\begin{bmatrix} 4 & -6 & 9 \\ -6 & 0 & -3 \\ 4 & 1 & 1 \end{bmatrix}_{3x3}$$

# of Rows = # of Columns

Diagonal Matrix

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 7 \end{bmatrix}_{3x3}$$

Non-Diagonal elements = 0

Identity (Unit) Matrix

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{3x3}$$

Diagonal elements =1 & Non-Diagonal elements = 0

Null (Zero) Matrix

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{3x3}$$

All elements = 0

Symmetric Matrix

$$\begin{bmatrix} 2 & 6 & -1 \\ 6 & 1 & 0 \\ -1 & 0 & 3 \end{bmatrix}_{3x3}$$

A' = A

Matrix Operations

Equality of Matrix

$$\begin{bmatrix} 4 & -6 & 9 \\ -6 & 0 & -3 \\ 4 & 1 & 1 \end{bmatrix}_{3x3} = \begin{bmatrix} 4 & -6 & 9 \\ -6 & 0 & -3 \\ 4 & 1 & 1 \end{bmatrix}_{3x3}$$

Respective elements of both matrix should be same i.e. $A_{ij} = B_{ij}$

Transpose of Matrix:

$$A = \begin{bmatrix} 4 & 2 & 9 \\ -6 & 5 & -3 \\ 4 & 5 & 1 \end{bmatrix}_{3x3}$$

$$A' = \begin{bmatrix} 4 & -6 & 4 \\ 2 & 5 & 5 \\ 9 & -3 & 1 \end{bmatrix}_{3x3}$$

Properties:

1. **(A')'=A**
2. **(kA)'=kA'**
3. **(A+B)'=A'+B'**
4. **(AB) '=B'A'**

Interchange Rows into Column (So that Columns into Rows)

Matrix Operations

Addition of Matrices:

$$\begin{bmatrix} 1 & 3 & -6 \\ -2 & 1 & 3 \\ 3 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 4 & 8 \\ 2 & 5 & 6 \\ 3 & 6 & 0 \end{bmatrix} = \begin{bmatrix} 1+2 & 3+4 & -6+8 \\ -2+2 & 1+5 & 3+6 \\ 3+3 & 0+6 & 1+0 \end{bmatrix} = \begin{bmatrix} 3 & 7 & 2 \\ 0 & 5 & 9 \\ 6 & 6 & 1 \end{bmatrix}$$

Add respective elements of both matrix

Properties:

1. **A+B = B+A (Communicative law)**
2. **(A+B)+C = A+(B+C) = (A+C)+B (Associative Law)**
3. **A+0 = 0+A = A**
4. **A+(-A) = 0**

Subtraction of Matrices:

$$\begin{bmatrix} 1 & 3 & -6 \\ -2 & 1 & 3 \\ 3 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 4 & 8 \\ 2 & 5 & 6 \\ 3 & 6 & 0 \end{bmatrix} = \begin{bmatrix} 1-2 & 3-4 & -6-8 \\ -2-2 & 1-5 & 3-6 \\ 3-3 & 0-6 & 1-0 \end{bmatrix} = \begin{bmatrix} -1 & -1 & -14 \\ -4 & -4 & -3 \\ 0 & -6 & 1 \end{bmatrix}$$

Subtract respective elements of both matrix

Matrix Operations

Multiplication of Matrices:

$$A=\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} 2\text{x}3 \qquad B=\begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} 3\text{x}2$$

Properties:

A(B+C) = AB+AC

(Distributive law)

(AB)C = A(BC) = (AC)B

(Associative Law)

IA = AI = A

A+(-A) = 0

$$AB=\begin{bmatrix} \boxed{(1*1)+(3*0)+(2*5)} & (1*3)+(3*1)+(2*2) \\ (4*1)+(0*0)+(1*5) & (4*3)+(0*1)+(1*2) \end{bmatrix}$$

$$=\begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix} 2\text{x}2$$

Matrix Operations

$$A=\begin{bmatrix} \boxed{1 & 3 & 6} \\ 2 & 1 & 3 \\ 3 & 0 & 1 \end{bmatrix}_{3\times3} \qquad B=\begin{bmatrix} 2 & 4 & 8 \\ 2 & 5 & 6 \\ 3 & 6 & 0 \end{bmatrix}_{3\times3}$$

$$AB=\begin{bmatrix} \boxed{(1*2)+(3*2)+(6*3)} & (1*4)+(3*5)+(6*6) & (1*8)+(3*6)+(6*0) \\ (2*2)+(1*2)+(3*3) & (2*4)+(1*5)+(3*6) & (2*8)+(1*6)+(3*0) \\ (3*2)+(0*2)+(1*3) & (3*4)+(0*5)+(1*6) & (3*8)+(0*6)+(1*0) \end{bmatrix}_{3\times3}$$

$$=\begin{bmatrix} 26 & 55 & 26 \\ 15 & 31 & 22 \\ 9 & 18 & 24 \end{bmatrix}_{3\times3}$$

Determinants

If $\quad ax + by = p \quad$ and

$\quad cx + dy = q \quad$ Then it can be represented as

$$Ax = b$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad x = \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } b = \begin{bmatrix} p \\ q \end{bmatrix}$$

If $(a\,d - b\,c \neq 0)$, then the system of linear equations has a unique solution.

The number$(a\,d - b\,c)$, which determines uniqueness of solution is associated with the matrix and is called the determinant of A and denoted by |A| or Δ or det.A.

Note:

1. Only square matrices have determinants
2. Expanding a determinant along any row or column gives same value.

Determinants

Properties:

- det(A)=det(A^T)
- If two rows (or columns) of A are equal, then det(A)=0.
- If a row (or column) of A consists entirely of 0, then det(A)=0.
- If B result from the matrix A by interchanging two rows (or columns) of A, then det(B)= - det(A)

Example: Evaluate the determinant $\Delta = \begin{vmatrix} 1 & 2 & 4 \\ -1 & 3 & 3 \\ 4 & 1 & 2 \end{vmatrix}$

So expanding along First Row (R1), we get

$$\Delta = \mathbf{1}\begin{vmatrix} \mathbf{3} & \mathbf{3} \\ \mathbf{1} & \mathbf{2} \end{vmatrix} - \mathbf{2}\begin{vmatrix} \mathbf{-1} & \mathbf{3} \\ \mathbf{4} & \mathbf{2} \end{vmatrix} + \mathbf{4}\begin{vmatrix} \mathbf{-1} & \mathbf{3} \\ \mathbf{4} & \mathbf{1} \end{vmatrix}$$

$$= \mathbf{1(3 \times 2 - 3 \times 1) - 2((-1) \times 2 - 3 \times 4) + 4((-1) \times 1 - 3 \times 4)}$$

$$= \mathbf{3 + 28 - 52}$$

$$= \mathbf{-21}$$

Linear Equations

Matrix Solution

If $\quad ax + by = p \quad and$

$\quad cx + dy = q$

Then it can be represented as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix}$$

$$Ax = b \qquad \text{where } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},\ X = \begin{bmatrix} x \\ y \end{bmatrix}\ and\ b = \begin{bmatrix} p \\ q \end{bmatrix}$$

$$A^{-1}Ax = A^{-1}b$$

$$Ix = A^{-1}b$$

$$x = A^{-1}b$$

Vector Projection

- Vector = Value + Direction
- Each point in a x,y coordinate system can be represented as a vector.
- Best Fit vector is a vector which best fits all the data points.
- We project all the vectors(corresponding to the data points to the best fit vector)
- Say we have 7 data points [(x1,y1), (x2,y2)…………….(x7,y7), we need to store, process 14 points.
- Vector representation of these 14 points will be u = u1, u2 & k1,k2,…..k7 (magnitudes of vectors) → We need 2 bits of u vector info and 7 values of respective vectors ie 9 points)
- This becomes valuable if we are dealing with millions of points in multiple dimensions.

Eigenvalues & Eigenvectors

- **If we apply matrix to eigen vector (multiply), it is like applying linear transformation resulting in λx (New Vector)**

$$Ax = \lambda x$$

- A is a square Matrix, x is a Eigen Vector (Value + Direction) not equal to zero.
- If we multiply a Matrix with Vector, the resultant output is proportional to the original Vector. (λ=-1,2)
- λ is a Eigen value
- Eigen Vector has a corresponding Eigen Value

Eigenvalues & Eigenvectors

- Example: Say a fish can be represented as a vector (having length ‘l’ and weight ‘w’. Each new fish will have a new l or w.
- Each fish in a 2 D space can be represented as a point or as a vector.
- If we apply a matrix A to all the vectors in the 2 D space, it will give us new vector .
- Linear Transformation of Matrix A with Eigen Vector, will transform to give new vectors with value λ (in the same direction of Eigen Vector)