Primer on Linear Algebra

Introduction to Vectors

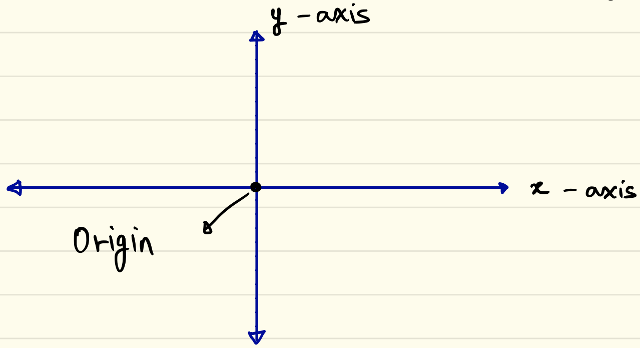
Perspectives of thinking about vectors :

Physics	Mathematics	Computer Science
- Direction and magnitude	- Generalized interpretation comprising of both Physics & CS perspective	- Ordered list

$$\vec{v}$$

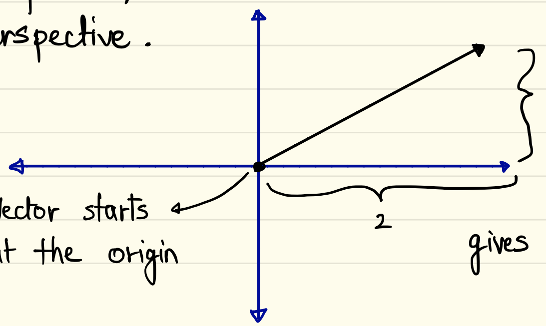
$$\begin{array}{cl} \left[\begin{array}{c} 2 \\ 1 \end{array}\right. & \begin{array}{l} x\text{-axis} \\ y\text{-axis} \end{array} \end{array}$$

- The place where the axes intersect is the origin.



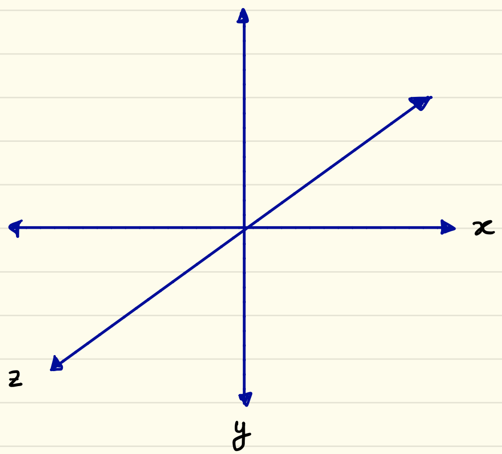
This system is the coordinate system.

- To obtain the CS perspective, we combine the coordinate system and physics perspective.

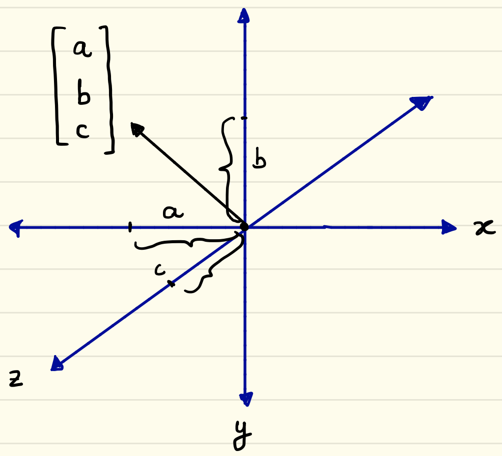


$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Extend to 3 - dimensions :



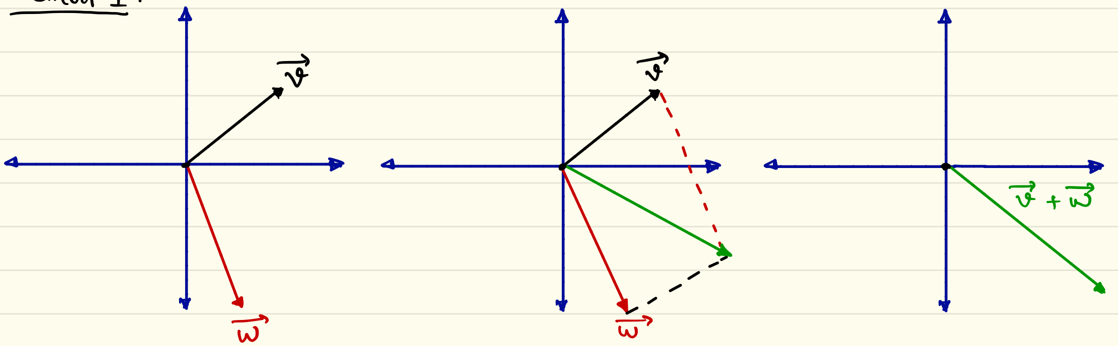
Any point $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is nothing but moving 'a' units in x-axis, moving 'b' units in y-axis and moving 'c' units in z-axis.



Now, we generalize the Physics + CS perspectives to get the math perspective.

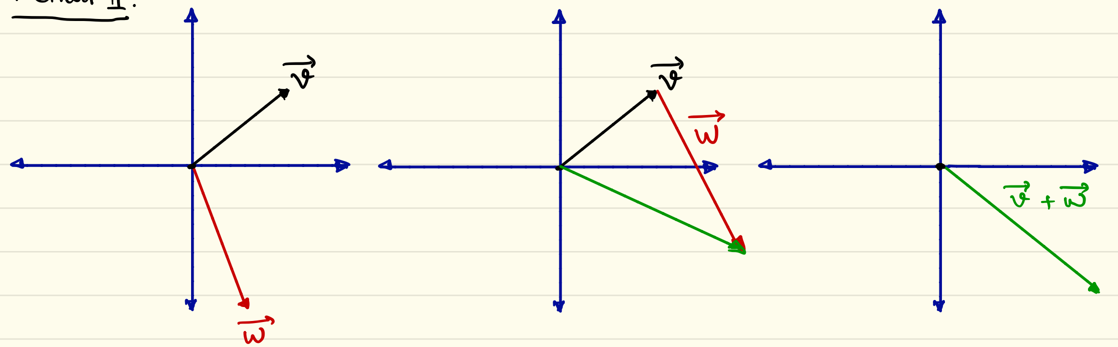
How to add two vectors $\vec{v}$ and $\vec{w}$?

Method I :



The addition process can also be visualized as :-

Method II :



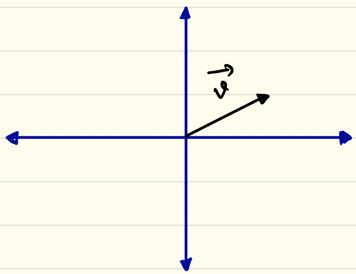
This means that shifting the origin is valid in vector algebra as long as the magnitude and direction remain the same.

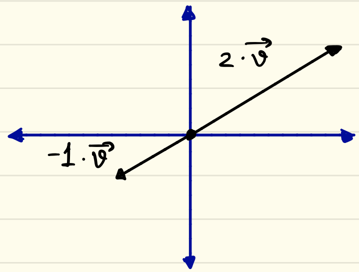
Numerically, if $\vec{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$, then,

$$\vec{v} + \vec{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$$

This is exactly what we achieved visually.

Now, let's talk about scaling vectors:





Numerically, $2 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ & $-1 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

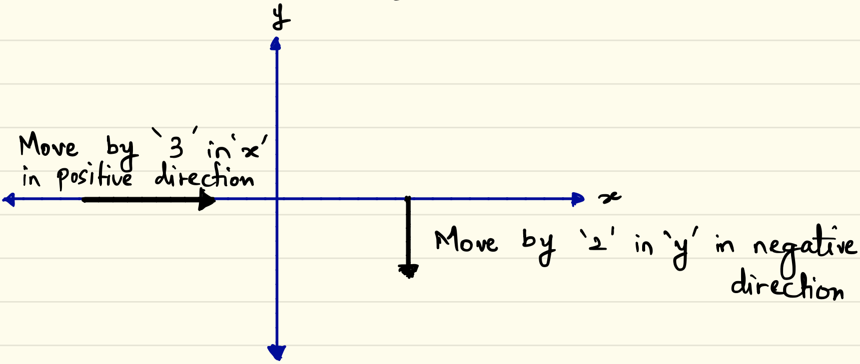
The above process is called scalar multiplication, where the vector is only scaled up or down without changing any other aspect of the vector.

Data Scientists can use this property to scale features up and down as per their utility.

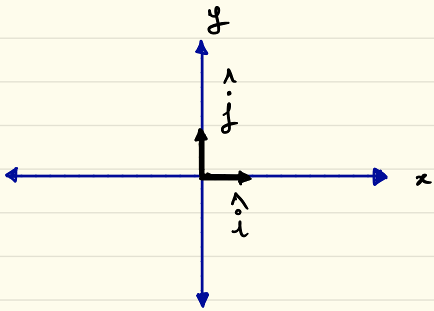
Basis, span, and linear dependence

Now, Let's think of each coordinate in $\begin{bmatrix} a \\ b \end{bmatrix}$ as a scalar.

for the case $a = 3$, $b = -2$, $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$ can be shown as:



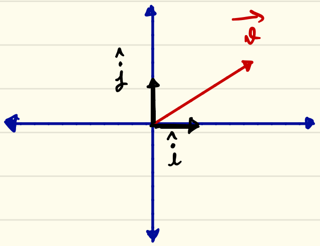
So, each vector can be thought of as a stretching and squishing in 'x' and 'y' direction. Let's define a general vector for that.



So, $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$ can now be thought of as $3\hat{i} + (-2)\hat{j}$.

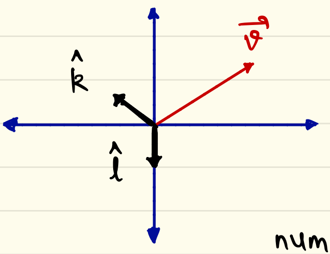
In the xy coordinate system, $\hat{i}$ and $\hat{j}$ can be called as 'Basis Vectors'.

In the previous case, we used $\hat{i}$ and $\hat{j}$ to be the basis vectors. However, we can choose some other basis vector too.



If $\hat{i}$ and $\hat{j}$ are basis vectors,

$$\vec{v} = 2\hat{i} + 1.5\hat{j}$$



If $\hat{k}$ and $\hat{l}$ are basis vectors,

$$\vec{v} = (-2)\hat{k} + (-1.5)\hat{l}$$

Therefore, when we represent any vector numerically, we assume basis vectors prior to defining them.

In general, the "span" of two vectors $\vec{v}$ and $\vec{w}$ is the set of all their possible linear combinations.

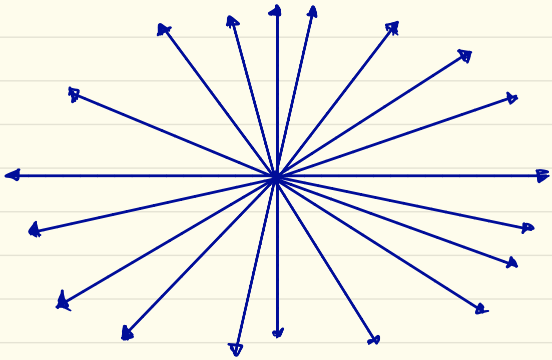
$\therefore$ $a\vec{v} + b\vec{w}$ where a & b can be real numbers.

In math notation:

$$\text{span}(\vec{v}, \vec{w}) = a\vec{v} + b\vec{w} \quad \underset{\text{(for all)}}{\forall}\ a, b \underset{\text{(Belongs to)}}{\in} \underset{\text{(Real)}}{\mathbb{R}}$$

Span of $\vec{v}$: 1 - dimensional line

Span of $\vec{v}$ and $\vec{w}$: 2 - dimensional plane



Span of $\vec{v}$, $\vec{w}$ and $\vec{z}$: 3 - dimensional space

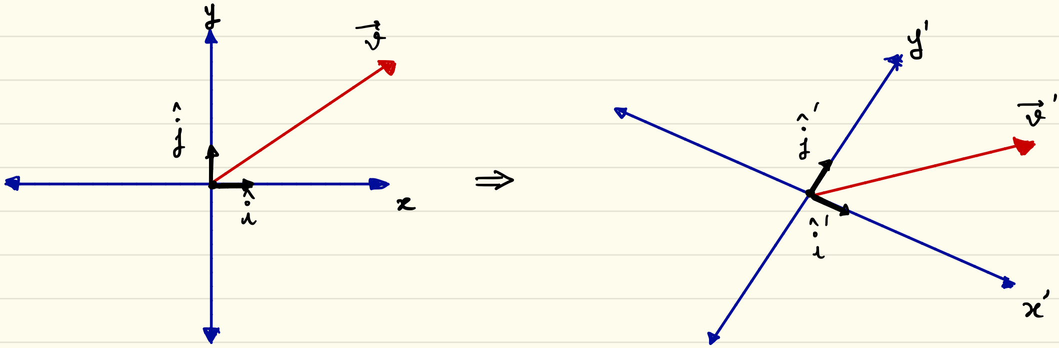
An interesting observation is, if $\vec{z} = a\vec{v} + b\vec{w}$, then the vector $\vec{z}$ does not add another dimension to the span of $\vec{v}$ and $\vec{w}$.

$\therefore$ $\vec{z} = a\vec{v} + b\vec{w}$ means 'z' is linearly dependent on $\vec{v}$ & $\vec{w}$

If adding $\vec{z}$ leads to an increase in dimension (2-d plane to 3-d space), then $\vec{z}$ is linearly independent to $\vec{v}$ & $\vec{w}$.

Technically, basis vectors are linearly independent vectors that span the entire space.

Linear Transformations, and Matrices



Linear transformation works when :

1. Origin remains fixed
2. Lines remain lines

Originally, $\vec{v} = a\,\hat{i} + b\,\hat{j}$

After transformation, $\vec{v}' = a\,\hat{i}' + b\,\hat{j}'$

If we keep track of $\hat{i} \rightarrow \hat{i}'$, and $\hat{j} \rightarrow \hat{j}'$, we can transform any vector without any other information.

if $\begin{bmatrix} \hat{i}' & \hat{j}' \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} & \begin{bmatrix} b \\ d \end{bmatrix} \end{bmatrix}$ (only two dimensions since its a 2-d plane)

then any general vector $\begin{bmatrix} x \\ y \end{bmatrix}$ lands on:

$$x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

Transformations can be chained, and it is important to look at matrix chaining in right-to-left manner, than traditional left-to-right.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} m & n \\ o & p \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

which implies,

$$\underset{M_3}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}\underset{M_2}{\begin{bmatrix} e & f \\ g & h \end{bmatrix}}\underset{M_1}{\begin{bmatrix} m & n \\ o & p \end{bmatrix}} = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix}$$

$$\longleftarrow$$

so, $\hat{i}$ & $\hat{j} \longrightarrow \hat{i}'$ & $\hat{j}'$ according to M_1

$\hat{i}'$ & $\hat{j}' \longrightarrow \hat{i}''$ & $\hat{j}''$ according to M_2

$\hat{i}''$ & $\hat{j}'' \longrightarrow \hat{i}'''$ & $\hat{j}'''$ according to M_3

Intuitive reason: it is just like $f(g(h(x)))$, where $h(x)$ is done first, then $g(h(x))$, and then $f(g(h(x)))$.

Note: In matrix multiplication:

Commutativity: $M_1 M_2 \neq M_2 M_1$

Associativity: $M_1(M_2 M_3) = (M_1 M_2)M_3$

Determinants

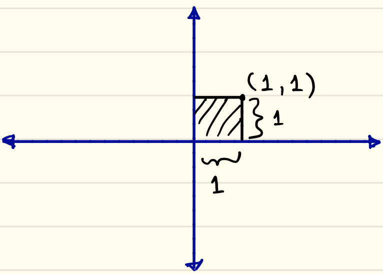
Consider a linear transformation of the 2-d space by $\begin{bmatrix} \hat{i}' & \hat{j}' \end{bmatrix}$ as $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$. Look at how the area changes.

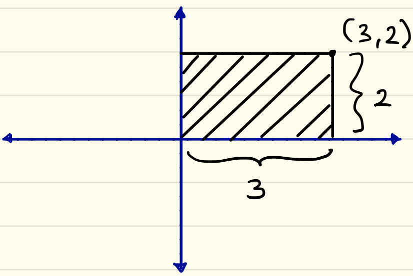
How does point $(1, 1)$ change?

By the transformation discussed previously:

$$1\begin{bmatrix} 3 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 3 + 1 \cdot 0 \\ 1 \cdot 0 + 1 \cdot 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Area under the curve increased (↑) 6 times.





By tracking changes to this 1 unit^2, we can talk about changes to any area on this 2-d plane, because any area can be approximated as a collection of unit squares.

So, $A' = 6 \cdot A$, where 6 is the determinant of the transformation.

Gaussian Elimination

How does Numpy, Pandas, and SciPy solve a system of linear equations?

A Data Scientist often is constrained within decision boundaries, and to identify a solution within the bounds of the linear system of equations, matrix algebra is used.

For eg; Let's say we encounter two equations on demand side and one from supply side as follows:

$$\left.\begin{aligned} x + y - z &= 9 \\ y + 3z &= 3 \end{aligned}\right\} \text{Demand side}$$

$$-x \quad - 2z = 2 \; \} \text{ Supply side}$$

This could perhaps be 3 different linear regressions which justifiably concurred with the hypotheses where null hypothesis was successfully rejected.

To solve these equations, "Gauss - Jordan Elimination" or "Gaussian Elimination" takes place within the libraries.

Scipy would represent the system as :-

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ -1 & 0 & -2 & 2 \end{array}\right]$$

$$\underbrace{\hphantom{1 \quad 1 \quad -1}}_{\text{LHS}} \quad \underbrace{\hphantom{9}}_{\text{RHS}}$$

LHS: Variables
RHS: Constraints

To solve this system, we need to somehow change the matrix's representation in a way that we are able to get 1's as the coefficient of each x, y, z so that we can isolate them.

This form is called "Row Echelon Form" of the matrix, which looks like:

$$\left[\begin{array}{ccc|c} 1 & a & b & d \\ 0 & 1 & c & e \\ 0 & 0 & 1 & f \end{array}\right]$$

Property 1 ✓

Property 2 ✓

Note:

Property 1: The leading entries in each of the rows of the matrix are '1', which isolates x, y, z.

Property 2: If a column contains a leading entry then all entries below that leading entry are 0.

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

Property 1 ✓

Property 2 X

Property 3 : In any two consecutive non-zero rows, the leading entry in the upper row occurs to the left of the leading entry in the lower row.

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

Property 1 ✓

Property 2 ✓

Property 3 ✓

Property 4 : All rows which consist entirely of zeroes appear at the bottom of the matrix.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Property 1 ✓

Property 2 ✓

Property 3 ✓

Property 4 ✓

Now that we understand Row Echelon form, let's solve the system of equations.

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ -1 & 0 & -2 & 2 \end{array}\right]$$

Need a zero here

Remember, whatever computation takes place in LHS has to take place in the RHS.

Step 1: Change row 3 to (row 3 + row 1)

$$R_1 + R_3 \to R_3$$

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ -1+1=0 & 0+1=1 & -2+(-1)=-3 & 2+9=11 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ -1 & 1 & -3 & 11 \end{array}\right]$$

Need a zero here

Step 2: $-1 \cdot R_2 + R_3 \to R_3$

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ -1\cdot(0)+0=0 & -1\cdot(1)+1=0 & -1(3)+(-3)=-6 & -1\cdot(3)+11=8 \end{array}\right]$$

$$= \left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ 0 & 0 & -6 & 8 \end{array}\right]$$

Need a one here

Step 3: $-\frac{1}{6} \cdot R_3 \rightarrow R_3$

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ -\frac{1}{6}\cdot(0) & -\frac{1}{6}\cdot(0) & -\frac{1}{6}\cdot(-6) & -\frac{1}{6}\cdot(8) \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 1 & -1 & 9 \\ 0 & 1 & 3 & 3 \\ 0 & 0 & 1 & -4/3 \end{array}\right]$$

So,

$$x + y - z = 9$$
$$y + 3z = 3$$
$$\boxed{z = -4/3}$$

$\therefore$ $y + 3\left(-\frac{4}{3}\right) = 3 \Rightarrow \boxed{y = 7}$

$x + 7 - \left(-\frac{4}{3}\right) = 9 \Rightarrow \boxed{x = 2/3}$

$$\boxed{\text{Solution: } \frac{2}{3},\ 7,\ -\frac{4}{3}}$$

Basically, through Gaussian Elimination, we eliminated dimensions which were linearly dependent.

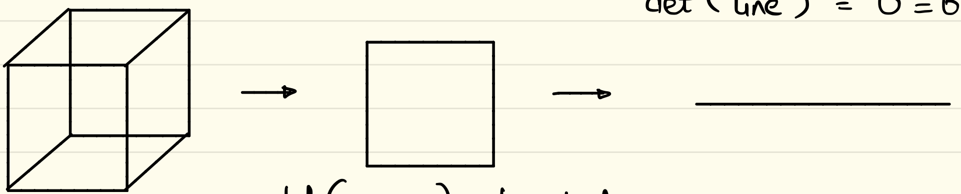
Data Scientists encounter 1000+ dimensions all the time. How cool is it to reduce the dimensions by squishing and rotating our 1000-dimensional space into 1-d!

Kinda like an image of a cube.

Assume cube of side '2'.

$\det(\text{cube}) = 8 = A$

$\det(\text{line}) = 0 = 0 \cdot A$



$\det(\text{square}) = 4 = \frac{1}{2} \cdot A$

3 - dimensions   2 - dimensions   1 - dimension

Representing $\underbrace{\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 3 \\ -1 & 0 & -2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\vec{u}} = \underbrace{\begin{bmatrix} 9 \\ 3 \\ 2 \end{bmatrix}}_{\vec{v}}$ in

mathematical notation:

$$A \vec{u} = \vec{v}$$

What we just solved through Gaussian elimination is:

$$\vec{u} = \underline{\underline{A^{-1}}} \vec{v}$$

So, to calculate A^{-1} all we need to do is solve:-

$$\left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ -1 & 0 & -2 & 0 & 0 & 1 \end{array}\right]$$

Zero here

We need this here, and then the RHS would be our A^{-1}.

Let's use Gaussian Elimination to do this.

Step 1: $R_1 + R_3 \to R_3$

$$\left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & 1 & -2 & 1 & 0 & 1 \end{array}\right]$$

zero here

Step 2: $-1 \cdot R_2 + R_3 \to R_3$

$$\left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & 0 & -6 & 1 & -1 & 1 \end{array}\right]$$

One here

Step 3 : $\frac{-1}{6} \cdot R_3 \to R_3$

$$\left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1/6 & 1/6 & -1/6 \end{array}\right]$$

zero here

Step 4 : $-1 \cdot R_2 + R_1 \to R_1$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -4 & 1 & -1 & 0 \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1/6 & 1/6 & -1/6 \end{array}\right]$$

zero here

Step 5 : $4 \cdot R_3 + R_1 \to R_1$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \left(-\frac{4}{6} + 1\right) & \left(\frac{4}{6} + (-1)\right) & \left(-\frac{4}{6} + 0\right) \\ 0 & 1 & 3 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1/6 & 1/6 & -1/6 \end{array}\right]$$

zero here

Step 6 : $-3R_3 + R_2 \to R_2$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/3 & -1/3 & 2/3 \\ 0 & 1 & 0 & \left(-\left(-\frac{3}{6}\right)+0\right) & \left(\left(-\frac{3}{6}\right)+1\right) & \left(-\left(\frac{-3}{6}\right)+0\right) \\ 0 & 0 & 1 & -1/6 & 1/6 & -1/6 \end{array}\right]$$

Voila!

Our solution is

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/3 & -1/3 & 2/3 \\ 0 & 1 & 0 & 1/2 & 1/2 & 1/2 \\ 0 & 0 & 1 & -1/6 & 1/6 & -1/6 \end{array}\right]$$

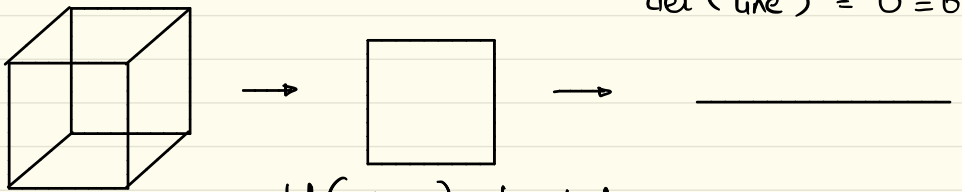
So,

$$\boxed{A^{-1} = \begin{bmatrix} 1/3 & -1/3 & 2/3 \\ 1/2 & 1/2 & 1/2 \\ -1/6 & 1/6 & -1/6 \end{bmatrix}}$$

Note: When the dimensions of a cube are squished like this:

$\det(\text{cube}) = 8 = A$

$\det(\text{line}) = 0 = 0 \cdot A$



$\det(\text{square}) = 4 = \frac{1}{2} \cdot A$

3 - dimensions   2 - dimensions   1 - dimension

We can come back to the cube from square as:

$$? \cdot (4) = 8 \Rightarrow \underline{\underline{? = 2}}$$

But we can't come back from a line:

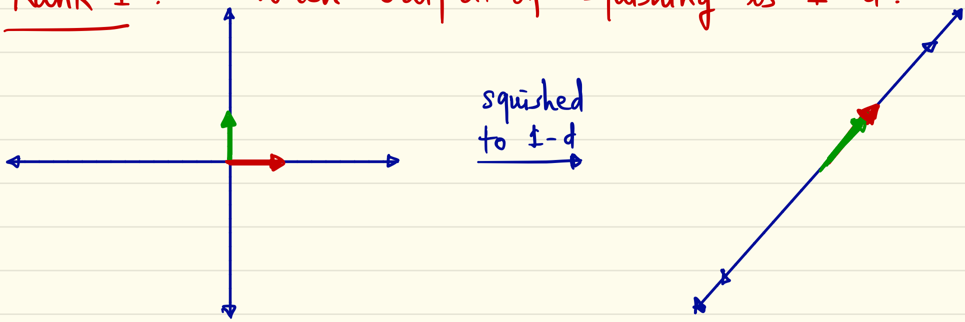
$$? \cdot (0) = 8 \Rightarrow ? \text{ can't be defined}$$

So, A^{-1} exists when $\det(A) \neq 0$.

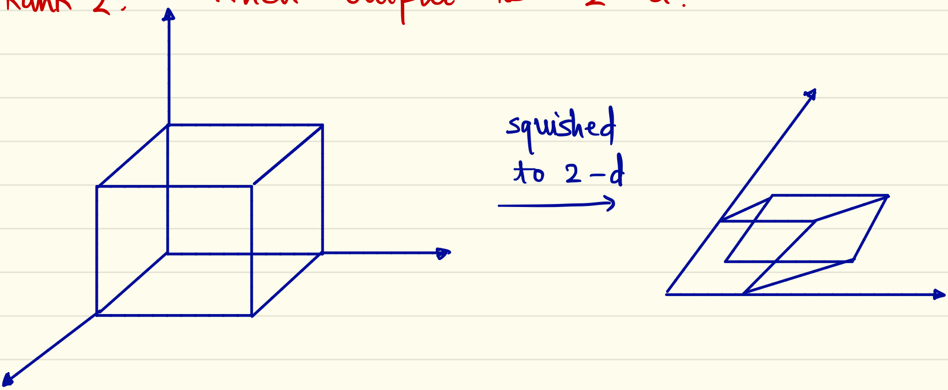
However, solutions may still exist for $\det(A) = 0$, but those solutions are stuck in the 1-dimension.

So, some squishes are squishier than others.

Rank 1 : When output of squishing is 1-d.



Rank 2 : When output is 2-d.



Rank is the "Number of dimensions in the output". More precisely, it is the "Number of dimensions in column space".

Note : The column space is the span of the column of the matrix. So, the column space of the column "age" for a data scientist working on demographics is all the possible space that is possible for "age" to occupy, both before and after preprocessing.