Tech I.S.

Course : Data Science

Module - I

# Maths for Data Science

*Content Curated By*

Abhinav J Navneet



January 23, 2023

# Chapter 1

# Introduction to Linear Algebra for Machine Learning



## 1.1 Why is Linear Algebra used in Data-Science and ML?

In Data Science and Machine Learning we work with data-sets. These data-sets are stored in form of rows and columns. Each column can be considered as an axis on a graph. Now in real world we only have 3 axis : x-axis, y-axis and z-axis.

What if the data-set have more than 3 columns ...like 100 columns!!! but we are given just 3-axis. Now what to do ?

Here Linear Algebra helps us. Linear Algebra provides us with tools that helps us work with any number of columns - meaning any number of axis.

There are two main tools that Linear Algebra provides:

1. Vectors

2. Vector spaces

Thus we can say that Vectors and Vector Spaces are the building blocks of Linear Algebra.

Note: The datasets having more than 3 columns are known as high-dimensional datasets.

### 1.1.1 How does vectors and vector spaces allows us to work with a higher dimensional dataset?

Vector can be though of as an object which is represented by - an ordered sequence of values.

Object1 = (a1, a2, a3, a4, a5)

Here Object1 is a vector which is represented by sequence of values - a1, a2, a3, a4 and a5

Now, this vector resides in an imaginary space that is beyond 3-dimensions - this imaginary space is called a Vector Space.

Here each value represents an imaginary axis. a1 is an axis, a2 is an axis and so on.

Size of a vector = no of values by which it is represented. For example: If we have a vector that is represented by 10 values then the size of vector is 10 and number of axis it has is also 10.



### 1.1.2 Visual representation of a Vector and Vector Space

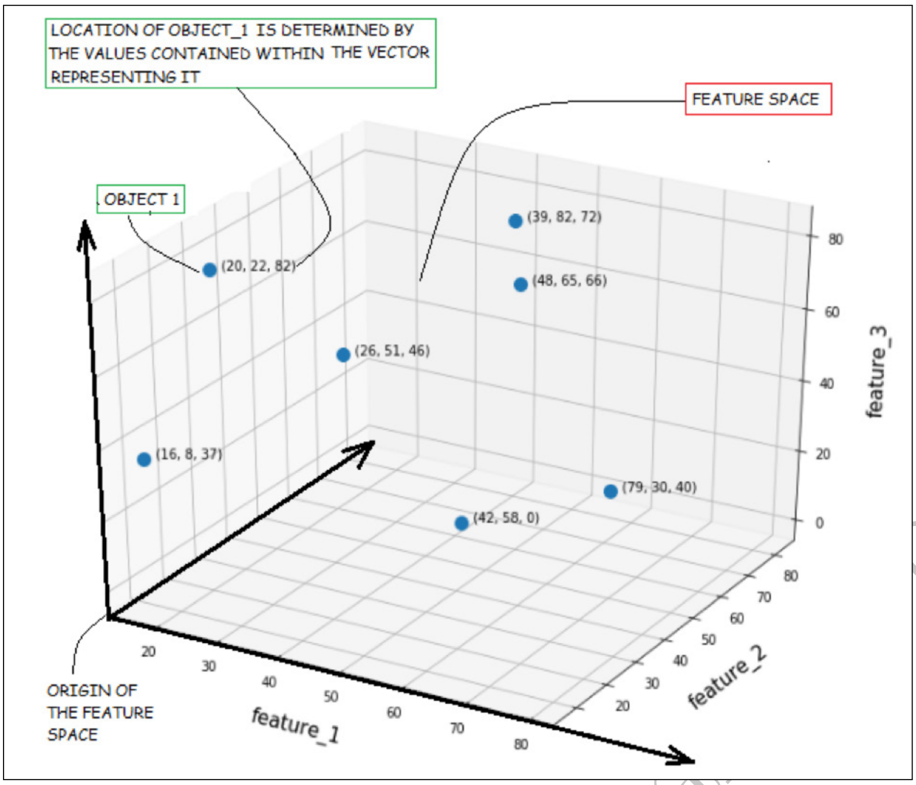
Vector Spaces can best be understood by visualizing 3 dimensional dataset.

Consider a dataset of consisting of 5 “objects”, described using 3 features (columns):

| | objects | feature_1 | feature_2 | feature_3 |
|---|---|---|---|---|
| 0 | 1 | 20 | 22 | 82 |
| 1 | 2 | 42 | 58 | 0 |
| 2 | 3 | 48 | 65 | 66 |
| 3 | 4 | 26 | 51 | 46 |
| 4 | 5 | 16 | 8 | 37 |

VECTOR REPRESENTING "OBJECT 1"

Now lets visualize the vector objects in 3-dimensions:




### 1.1.3 How is a vector represented mathematically?

Vectors are mathematically represented as shown below.

The alphabet representing the vector is in bold small letters and it has a “dash” crowning it:

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

Elements of the vector are represented as a "column" within a square bracket.

## 1.2 Introduction to Vector Operations

### 1.2.1 What do we understand by Vector Operations?

Consider that there are two vectors - v1 and v2. Now, how these 2 vectors will interact with each other can be understood by vector operations.

There are many vector operations in mathematics but the ones that are most relevant for the purpose of DataScience & ML are:

1. Vector Addition ( which is also known as vector combination)

2. Vector Scaling (which is also known as vector multiplication)

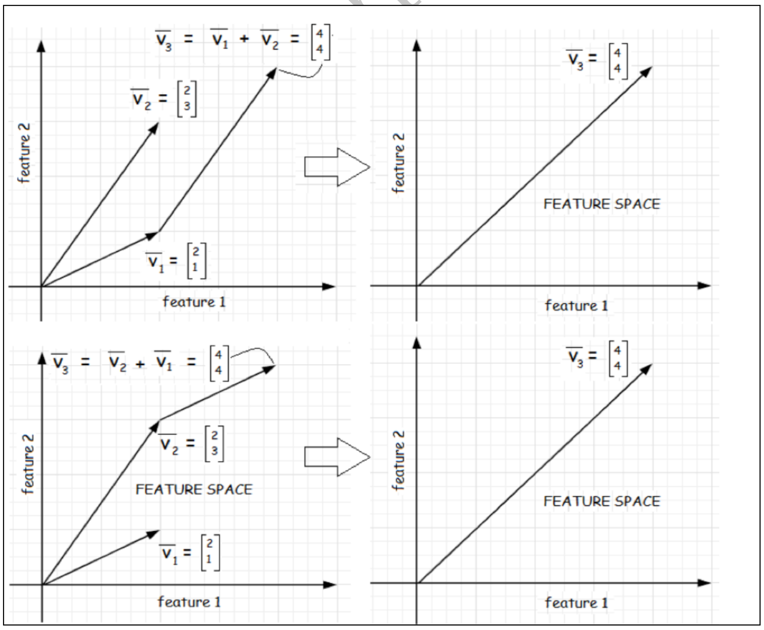
3. Vector Dot Product

**What is Vector Addition?**

Consider that there are two vectors v1 and v2. Each of these vectors have their own values and thus locations in the imaginary space.

Now vector addition is the operation that gives us a new vector v3 which is resultant when we combine/add v1's and v2's values and locations.

Observe the representation given below:



The above mentioned vector addition is mathematically expressed as:

$$\overline{V_3} = \overline{V_1} + \overline{V_2} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

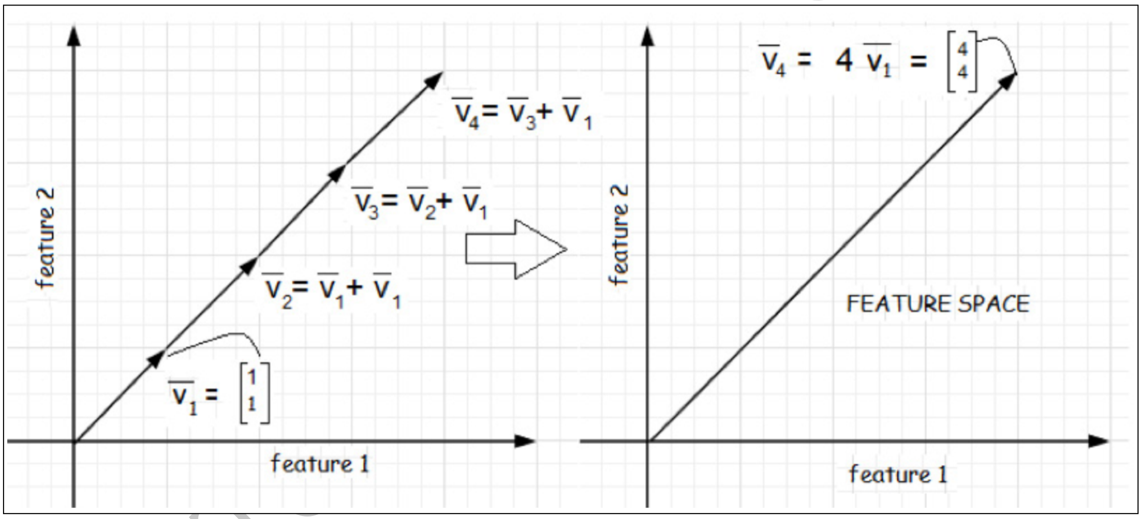
**What is Vector Scaling or Vector Multiplication?**

Consider we have a single vector v1.

Now Vector Scaling or Vector Multiplication will give us a new vector v4 which is resultant of adding v1 n number of times.

Consider the image shown below, the location represented by vector v4 is the resultant location we would reach if we sequentially moved/added 4 times to the "relative location" represented by vector v1.






The above mentioned Vector scaling is mathematically expressed as:

$$\overline{V_4} = 4\overline{V_1} = 4\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

In General:

$$a\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} a\,f_1 \\ a\,f_2 \end{bmatrix}$$

**What is Vector Dot product?**

The dot product of 2 vectors v1 & v2, belonging to the same feature space is mathematically defined as:

$$\bar{v}_1 \cdot \bar{v}_2 = \begin{bmatrix} f_{1,1} \\ f_{2,1} \end{bmatrix} \cdot \begin{bmatrix} f_{1,2} \\ f_{2,2} \end{bmatrix} = f_{1,1}\, f_{1,2} + f_{2,1}\, f_{2,2} \text{ —— For 2 dimensional feature spaces}$$

Mathematical properties of a dot product:

$$\bar{v}_1 \cdot \bar{v}_2 = \bar{v}_2 \cdot \bar{v}_1 \text{ —— COMMUTATIVITY}$$

$$(\bar{v}_1 + \bar{v}_2) \cdot \bar{v}_3 = \bar{v}_1 \cdot \bar{v}_3 + \bar{v}_2 \cdot \bar{v}_3 \text{ —— DISTRIBUTIVITY}$$

$$a\,\bar{v}_1 \cdot \bar{v}_2 = a\,(\bar{v}_1 \cdot \bar{v}_2) \text{ —— ASSOCIATIVITY}$$



## 1.3 Introduction to commonly used Vector concepts

### 1.3.1 What is Co-linearity and Linear Independence for vectors?

Two vectors v1 & v2 are said to be collinear if one of them can be expressed as a "scaled" version of the other.

Conversely, if vectors v1 & v2 are non collinear, they are said to be Linearly Independent of each other.

Vectors $\bar{V}_1$ & $\bar{V}_2$ are said to be collinear if:

$$\bar{V}_1 = a\,\bar{V}_2$$

### 1.3.2 What do you mean by distance between two vectors?

The distance between two vectors would mean the distance between the locations represented by the two vectors.

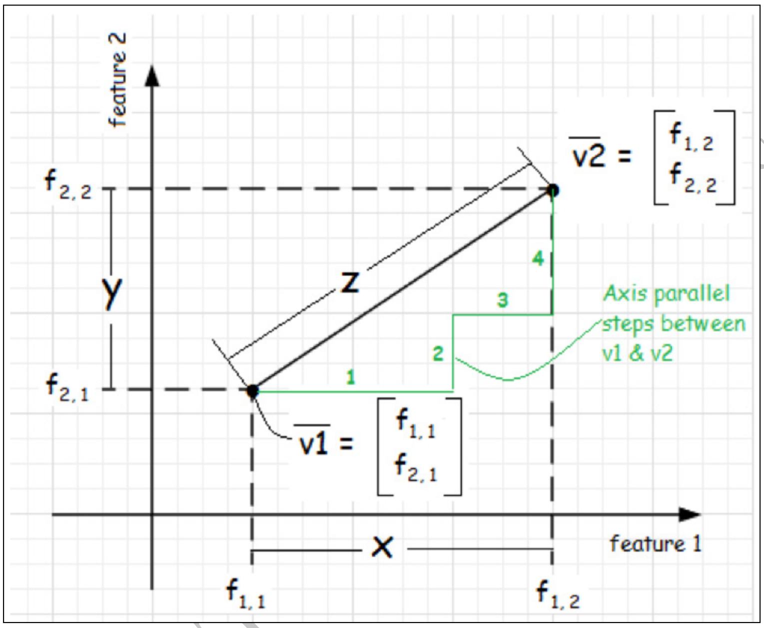
Type of Distances commonly used:

1. Euclidean Distance

2. Minkowski's Distance & Norms

3. Manhattan Distance

The concept of distance is best understood with reference to 2 dimensional feature spaces, So for understanding the above distances consider the following representation:



**What is Euclidean Distance?**

Euclidean Distance is a measure of distance based on the Pythagorean theorem.

With reference to the image above, the Euclidean distance is given by the value of “z”, which is equal to the square root of the sum of the squares of “x” & “y”. Euclidean distance is expressed mathematically as shown below:

$$\|\overline{v1} - \overline{v2}\|_2 = \left( |f_{1,2} - f_{1,1}|^2 + |f_{2,2} - f_{2,1}|^2 \right)^{1/2} \text{ — For 2 dimensional feature spaces}$$

$$\underbrace{\|\overline{v1} - \overline{v2}\|_2}_{\text{*Norm 2}} = \left( \sum_{i=1}^{n} |f_{i,2} - f_{i,1}|^2 \right)^{1/2} \text{ — Generalized for "n" dimensional feature space}$$

Special brackets denote absolute value

**What is Minkowski’s Distance? & Norm ?**

The Minkowski’s distance is a generalization of the Pythagorean theorem with respect to the exponent term used within its formula:

Here, Z is also called Norm of p value.

$$z^p = x^p + y^p$$

The Minkowski’s distance is expressed mathematically as shown below:

$$\|\overline{v2} - \overline{v1}\|_p = \left( |f_{1,2} - f_{1,1}|^p + |f_{2,2} - f_{2,1}|^p \right)^{1/p} \text{ — For 2 dimensional feature spaces}$$

$$\|\overline{v2} - \overline{v1}\|_p = \left( \sum_{i=1}^{n} |f_{i,2} - f_{i,1}|^p \right)^{1/p} \text{ — Generalized for "n" dimensional feature space}$$

**What do you understand by Magnitude of a vector?**

The Magnitude of a vector is used to describe the Euclidean distance of that vector, from the origin of its feature space.

Thus the magnitude of a vector is mathematically expressed as:

$$\|\bar{v}\|_2 = \left( |f_1|^2 + |f_2|^2 \right)^{1/2} \text{ —— For 2 dimensional feature spaces}$$

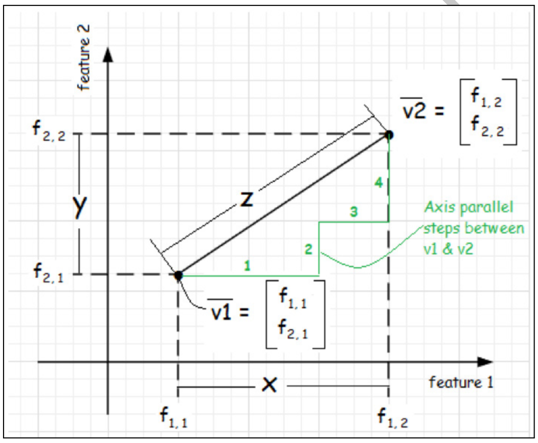
$$\|\bar{v}\|_2 = \left( \sum_{i=1}^{n} |f_i|^2 \right)^{1/2} \text{ —— Generalized for "n" dimensional feature space}$$

**What is Manhattan Distance?**

Manhattan is a grid based distance measure.

Manhattan distance is the distance between vectors v1 & v2 if we only take "axis-parallel" approach/steps (ie: Any single step has to be parallel to at least one axis of the feature space).

In other words, the Manhattan distance is the "Norm 1" interpretation of distance and it is mathematically expressed as shown below:




$$\|\overline{v2} - \overline{v1}\|_1 = |f_{1,2} - f_{1,1}| + |f_{2,2} - f_{2,1}| \text{ —— For 2 dimensional feature spaces}$$

$$\underbrace{\|\overline{v2} - \overline{v1}\|_1}_{\text{Norm 1}} = \sum_{i=1}^{n} |f_{i,2} - f_{i,1}| \text{ —— Generalized for "n" dimensional feature space}$$

# Chapter 2

# Introduction to Matrices

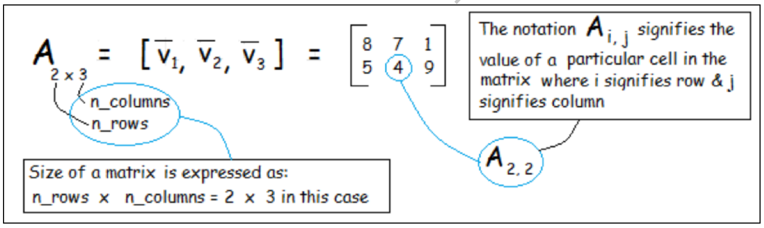
## 2.1 What is a Matrix ??



A matrix is a collection of vectors that belongs to the same feature space.

The vectors are enclosed within a square bracket, each of these vectors represent a column of the matrix. Now from the point of view of a dataset we can say that - each 'column' within the matrix represents a data-object (ie: vector) and each 'row' represents the individual features of the data-objects.

Shown below is the mathematical representation of a matrix containing three 2D vectors:

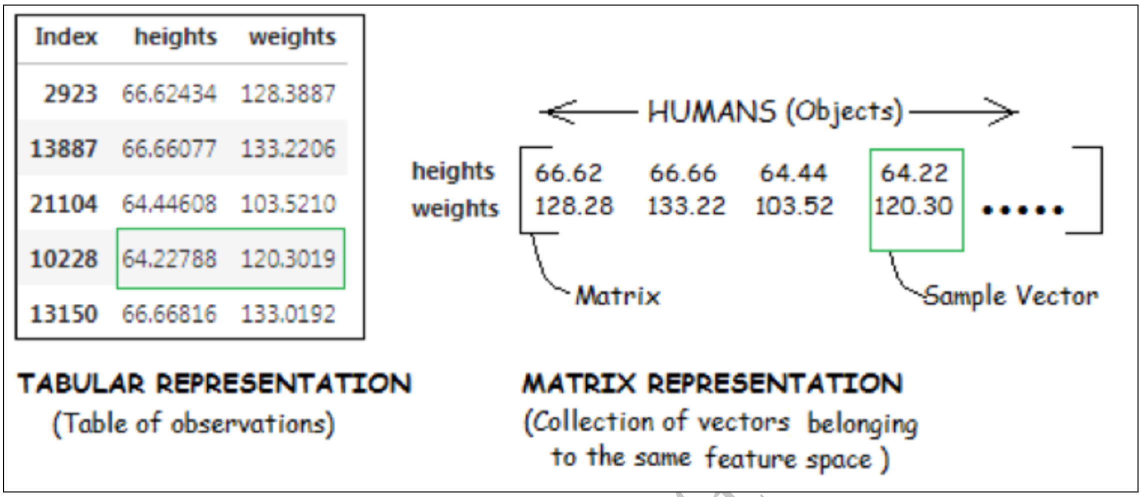



Note that the alphabet representing the matrix is in bold capitals and has a subscript expressing the matrix dimensions (i.e:size).

### 2.1.1 Example of a matrix in real life:

Consider a table(dataset) consisting of two columns - Heights and Weights. Here Heights and Weights are Vectors.

Below you can see the representation of this table in form of a matrix:




### 2.1.2 Why do we use matrix?

Notice the difference in how data is represented in the " generic" Tabular format and the "formal" matrix format.

The tabular format is more readable, but the matrix representation is more suited for understanding Linear Algebraic concepts like matrix transformations. Matrix transformations are one of the backbone of Machine Learning Algorithms.

### 2.1.3 What is 'Transpose of a matrix' ?

When we transpose a matrix, we get a new matrix such that rows becomes the columns and columns becomes the rows. The transpose of a matrix is expressed as shown below:

Indicates Transpose of matrix A

$$A = \begin{bmatrix} 8 & 7 & 1 \\ 5 & 4 & 9 \end{bmatrix} \qquad A^T = \begin{bmatrix} 8 & 5 \\ 7 & 4 \\ 1 & 9 \end{bmatrix} \qquad \text{So, } (A^T)^T = A$$

### 2.1.4 What is 'Diagonal of a matrix'?

The diagonal of a matrix consists of those elements, where the row number & column number are same. Shown below are 3 example matrices with their diagonals marked:

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

For all $A_{i,j}$ $B_{i,j}$ & $C_{i,j}$ in above matrices, the elements with i = j form the diagonal

### 2.1.5 What are Square and Symmetric matrices?

Square matrices are those matrices which have the same number of rows and columns. i.e i=j , where i and j are matrices

Symmetric matrices are square matrices which have the same elements, in the same (reflective) order, on either sides of their diagonal.

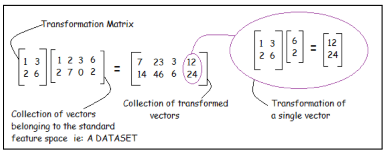
In other words, if the transpose of a matrix results in the same matrix, then that matrix is a symmetric matrix. The example below demonstrates this:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

### 2.3.1 What is a Diagonal Matrix and Identity Matrix?

Diagonal Matrices are matrices in which all elements except for the diagonal have zero values.

An identity matrix is a diagonal matrix with only unit values. Any vector transformed by an identity matrix results is itself (i.e: An identity matrix does not produce any "transformation"), as shown below:




### 2.3.2 What is Matrix Multiplication?

Matrix Multiplication is nothing but - Matrix transformations of a group of vectors.

In other words, multiple vectors(represented as a matrix on the right) are simultaneously transformed by a single Transformation Matrix on the left.

Note: The Transformation Matrix could thus be viewed as mapping function between two different feature spaces.

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

### 2.3.3 What is Matrix Inverse?

Matrix Inverse can be viewed as Inverse of Transformation matrix.

Up till now we were transforming a vector v in standard space S1 to fit alternate feature S2. Now what if we want to reverse this, i.e. go from S2 back to S1. For this we use Matrix Inverse.

In other words, given a transformation matrix that transforms vectors from standard feature space to an alternative feature space, the matrix inverse of the transformation matrix is a unique transformation matrix that transforms vectors from the alternative feature space to the standard feature space.

For a given matrix A, the inverse of a matrix A is represented as:

Inverse of matrix $A = A^{-1}$

If $AV = V'$, then $A^{-1}V' = V$