

Road to Data Science in 50 Days - Day 5 ¶

Statistics for Data Science

Statistics for Data Science



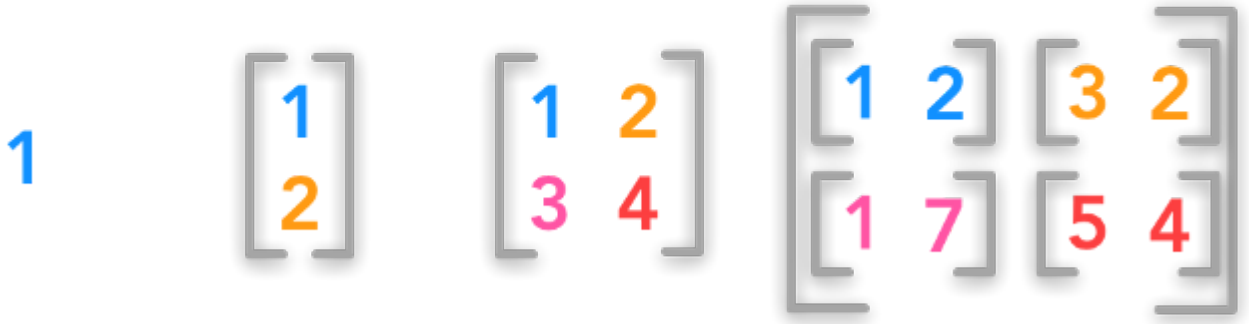
There's 2 more part of math that is usually used in Data Science i.e. Linear Algebra and Bayesian Statistics.

After thourougly going through various articles on Linear Algebra, i have realized that linear algebra would be pretty difficult to learn from written blogs and it needs video explanations. So we will get introduced to the basics of Linear Algebra and suggest some videos to go through in the end.

Linear Algebra

In linear algebra, you will often hear the terms vectors, matrix or tensors in deep learning. Let's get introduced to these terms:

Scalar Vector Matrix Tensor



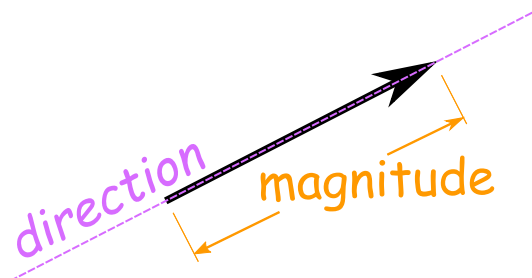
1. Scalar:

A scalar is usually said to be a physical quantity that only has magnitude, possibly a sign, and no other characteristics

2. Vector:



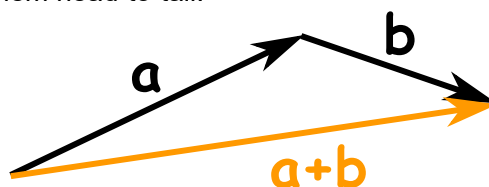
A vector has magnitude (size) and direction



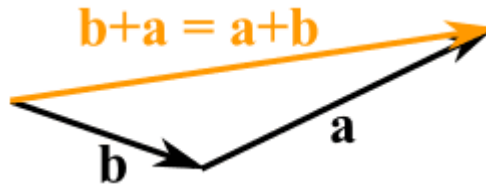
The length of the line shows its magnitude and the arrowhead points in the direction.

Addition of 2 Vectors

We can add two vectors by joining them head-to-tail:

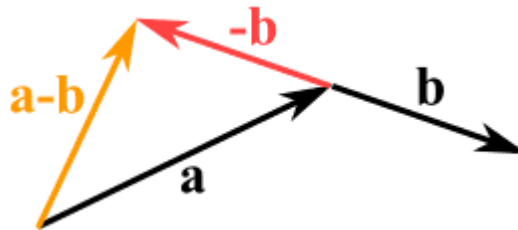


And it doesn't matter which order we add them, we get the same result:



We can also subtract one vector from another:

- first we reverse the direction of the vector we want to subtract,
- then add them as usual:



Magnitude of a Vector

The magnitude of a vector is shown by two vertical bars on either side of the vector:

$$|a|$$

OR it can be written with double vertical bars (so as not to confuse it with absolute value):

$$||a||$$

We use Pythagoras' theorem to calculate it:

$$|a| = \sqrt{x^2 + y^2}$$

Example:

what is the magnitude of the vector $b = (6, 8)$?

$$|b| = \sqrt{6^2 + 8^2} = \sqrt{36 + 64} = \sqrt{100} = 10$$

A vector with magnitude 1 is called a Unit Vector (<https://www.mathsisfun.com/algebra/vector-unit.html>).

**To learn more about vectors in details, refer: <https://www.mathsisfun.com/algebra/vectors.html>
(<https://www.mathsisfun.com/algebra/vectors.html>)

3. A Matrix is an array of numbers

$$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}$$

(This one has 2 Rows and 3 Columns)

There are many things we can do with them ...

Adding

To add two matrices: add the numbers in the matching positions:

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

3+4=7

These are the calculations:

$$3+4=7 \quad 8+0=8$$

$$4+1=5 \quad 6-9=-3$$

The two matrices must be the same size, i.e. the rows must match in size, and the columns must match in size.

Negative

The negative of a matrix is also simple:

$$-\begin{bmatrix} 2 & -4 \\ 7 & 10 \end{bmatrix} = \begin{bmatrix} -2 & 4 \\ -7 & -10 \end{bmatrix}$$

-(2)=-2

Matrix Negative

These are the calculations:

$$-(2)=-2 \quad -(-4)=+4$$

$$-(7)=-7 \quad -(10)=-10$$

Subtracting

To subtract two matrices: subtract the numbers in the matching positions:

$$\begin{bmatrix} 3 & 8 \\ 4 & 6 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} -1 & 8 \\ 3 & 15 \end{bmatrix}$$

$3-4=-1$

Matrix Subtraction

These are the calculations:

$$3-4=-1 \quad 8-0=8$$

$$4-1=3 \quad 6-(-9)=15$$

Multiply by a Constant

We can multiply a matrix by a constant (the value 2 in this case):

$$2 \times \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 2 & -18 \end{bmatrix}$$

$2 \times 4 = 8$

These are the calculations:

$$2 \times 4 = 8 \quad 2 \times 0 = 0$$

$$2 \times 1 = 2 \quad 2 \times -9 = -18$$

We call the constant a scalar, so officially this is called "**scalar multiplication**".

Multiplying by Another Matrix

$$\begin{bmatrix} A & B \\ C & D \\ E & F \end{bmatrix} \times \begin{bmatrix} G \\ H \end{bmatrix} = \begin{bmatrix} A \times G + B \times H \\ C \times G + D \times H \\ E \times G + F \times H \end{bmatrix}$$

To multiply two matrices together is a bit more difficult ... read [Multiplying Matrices](https://www.mathsisfun.com/algebra/matrix-multiplying.html) (<https://www.mathsisfun.com/algebra/matrix-multiplying.html>) to learn how.

Dividing

And what about division? Well we don't actually divide matrices, we do it this way:

$$A/B = A \times (1/B) = A \times B^{-1}$$

where B^{-1} means the "inverse" of B .

So we don't divide, instead we multiply by an inverse.

And there are special ways to find the Inverse, learn more at [Inverse of a Matrix](https://www.mathsisfun.com/algebra/matrix-inverse.html) (<https://www.mathsisfun.com/algebra/matrix-inverse.html>).

Transposing

To "transpose" a matrix, swap the rows and columns.

$$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}^T = \begin{bmatrix} 6 & 1 \\ 4 & -9 \\ 24 & 8 \end{bmatrix}$$

We put a "T" in the top right-hand corner to mean transpose

Types of Matrix

$$\begin{bmatrix} 7 & 6 & 4 \\ 4 & 2 & -2 \\ 3 & 0 & 9 \end{bmatrix}$$

There are various types of matrix such as:

1. Diagonal Matrix
2. Inverse Matrix
3. Square Matrix
4. Identity Matrix
5. Scalar Matrix
6. Triangular Matrix
7. Symmetric
8. Hermitian

To learn about the types of matrix in details:, refer <https://www.mathsisfun.com/algebra/matrix-types.html> (<https://www.mathsisfun.com/algebra/matrix-types.html>)

Notation

A matrix is usually shown by a capital letter (such as A, or B)

Each entry (or "element") is shown by a lower case letter with a "subscript" of row,column:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}$$

Rows and Columns So which is the row and which is the column?

- Rows go left-right
- Columns go up-down

To remember that rows come before columns use the word "arc":

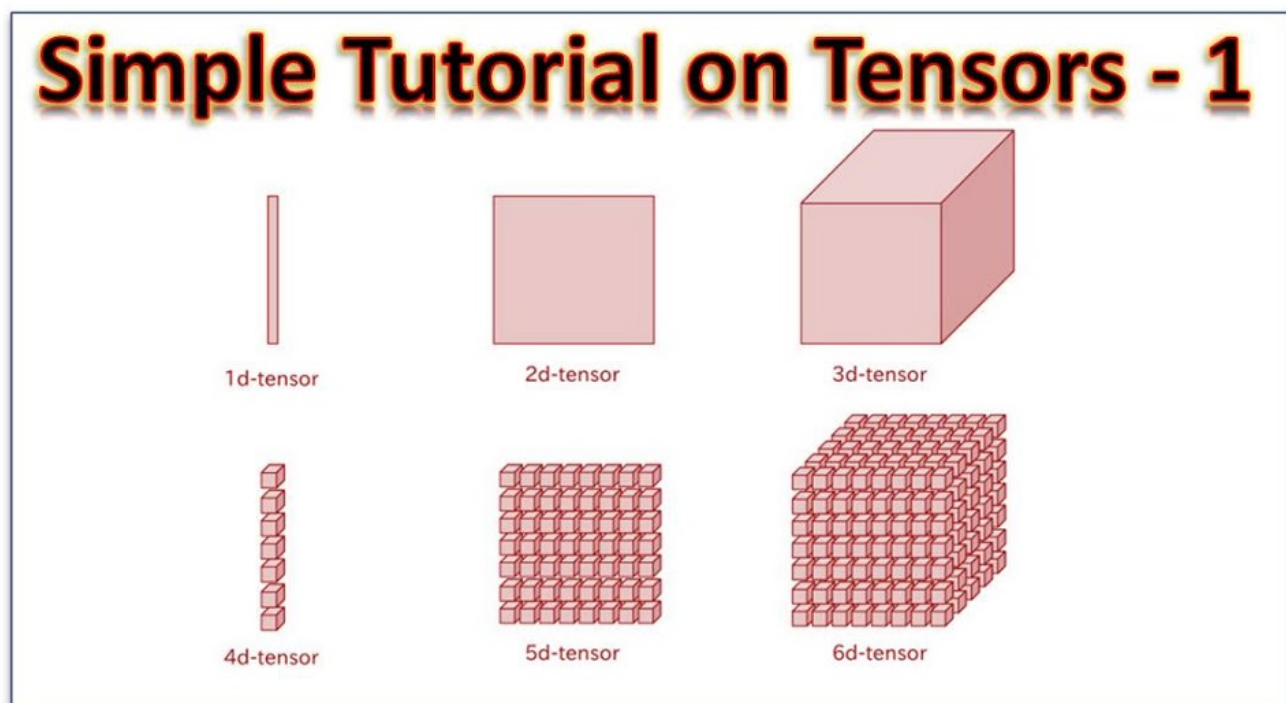
A_{r,c}

4. Tensors

Well tensors are a bit too much to learn at the basic, i would suggest learn and practise more and more on vectors and matrices calculations before starting with Tensors.

Tensors can be defined in multiple ways:

1. **Abstract way:** they are mathematical objects, that has certain properties, and transforms in a particular way.
1. **Algebraic way:** they are a generalization of the usual 2D matrices. That is, one can think of them as a transformation rule for objects/coordinates in some n-dimensional (vector) space.
2. **Geometrical way:** they are a collection of tangent covectors at one point of some manifold, mixed with another collection of linear functionals associated with the same point, but located in dual space (this explanation is rough), even so, this is the most general and precise way to think about tensors.
1. **Physical way:** this, in my opinion, is the best way to visualize them, it is like the algebraic one, has the same issues, but still more general and intuitive.



To learn more about Tensors, refer this really fun yet informative video tutorial: <https://www.youtube.com/watch?v=bpG3ggDM80w> (<https://www.youtube.com/watch?v=bpG3ggDM80w>).

Eigen Values and Eigen Vectors

Eigenvectors and eigenvalues live in the heart of the data science field. It's a must-know topic for anyone who wants to understand machine learning in-depth.

Eigenvalues and eigenvectors form the basics of computing and mathematics. They are heavily used by scientists.

The below explanations are going to be from this very nicely written article by Farhad Malik on Medium:

<https://medium.com/fintechexplained/what-are-eigenvalues-and-eigenvectors-a-must-know-concept-for-machine-learning-80d0fd330e47> (<https://medium.com/fintechexplained/what-are-eigenvalues-and-eigenvectors-a-must-know-concept-for-machine-learning-80d0fd330e47>)

1. Eigenvectors and Eigenvalues Introduction

1.1 What Is An Eigenvector?

I would like to explain this concept in a way that we can easily understand it. For the sake of simplicity, let's consider that we live in a two-dimensional world.

- Alex's house is located at coordinates $[10,10]$ ($x=10$ and $y=10$). Let's refer to it as vector A.
- Furthermore, his friend Bob lives in a house with coordinates $[20,20]$ ($x=20$ and $y=20$). I will refer to it as vector B.

If Alex wants to meet Bob at his place then Alex would have to travel +10 points on the x-axis and +10 points on the y-axis. This movement and direction can be represented as a two-dimensional vector $[10,10]$. Let's refer to it as vector C.

We can see that vector A to B are related because vector B can be achieved by scaling (multiplying) the vector A by 2. This is because $2 \times [10,10] = [20,20]$. This is the address of Bob. Vector C also represents the movement for A to reach B.

The key to note is that a vector can contain the magnitude and direction of a movement. So far so good! We learned from the introduction above that large set of data can be represented as a matrix and we need to somehow compress the columns of the sparse matrix to speed up our calculations. Plus if we multiply a matrix by a vector then we achieve a new vector. The multiplication of a matrix by a vector is known as transformation matrices.

The new vector can be considered to be in two forms:

1. Sometimes, the new transformed vector is just a scaled form of the original vector. This means that the new vector can be re-calculated by simply multiplying a scalar (number) to the original vector; just as in the example of vector A and B above.
2. And other times, the transformed vector has no direct scalar relationship with the original vector which we used to multiply to the matrix.

Therefore, if our input is a large sparse matrix M then we can find a vector o that can replace the matrix M. The criteria is that the product of matrix M and vector o should be the product of vector o and a scalar n:

$$M * o = n * o$$

This means that a matrix M and a vector o can be replaced by a scalar n and a vector o . In this instance, o is the eigenvector and n is the eigenvalue and our target is to find o and n .

Therefore an eigenvector is a vector that does not change when a transformation is applied to it, except that it becomes a scaled version of the original vector.

Eigenvectors can help us calculating an approximation of a large matrix as a smaller vector. There are many other uses which I will explain later on in the article.

Eigenvectors are used to make linear transformation understandable. Think of eigenvectors as stretching/compressing an X-Y line chart without changing their direction.

1.2 What is an Eigenvalue?

Eigenvalue—The scalar that is used to transform (stretch) an Eigenvector.

2. Where are Eigenvectors and Eigenvalues used?

There are multiple uses of eigenvalues and eigenvectors:

1. Eigenvalues and Eigenvectors have their importance in linear differential equations where you want to find a rate of change or when you want to maintain relationships between two variables.

Think of eigenvalues and eigenvectors as providing summary of a large matrix

1. We can represent a large set of information in a matrix. Performing computations on a large matrix is a very slow process. To elaborate, one of the key methodologies to improve efficiency in computationally intensive tasks is to reduce the dimensions after ensuring most of the key information is maintained. Hence, one eigenvalue and eigenvector are used to capture key information that is stored in a large matrix. This technique can also be used to improve the performance of data churning components.
2. Component analysis is one of the key strategies that is utilised to reduce dimension space without losing valuable information. The core of component analysis (PCA) is built on the concept of eigenvalues and eigenvectors. The concept revolves around computing eigenvectors and eigenvalues of the covariance matrix of the features.
3. Additionally, eigenvectors and eigenvalues are used in facial recognition techniques such as EigenFaces.
4. They are used to reduce dimension space. The technique of Eigenvectors and Eigenvalues are used to compress the data. As mentioned above, many algorithms such as PCA rely on eigenvalues and eigenvectors to reduce the dimensions
5. Eigenvalues are also used in regularisation and they can be used to prevent overfitting.

Eigenvectors and eigenvalues are used to reduce noise in data. They can help us improve efficiency in computationally intensive tasks. They also eliminate features that have a strong correlation between them and also help in reducing over-fitting.

1. We can also use eigenvector to rank items in a dataset. They are heavily used in search engines and calculus.

Having said that, it can be slow to compute eigenvectors and eigenvalues. The computation is $O(n^3)$

It is now apparent that Eigenvalues and Eigenvectors are one of core concepts to understand in data science.

3. Calculating Eigenvectors and Eigenvalues

Although we don't have to calculate the Eigenvalues and Eigenvectors by hand every time but it is important to understand the inner workings to be able to confidently use the algorithms.

4.1 Let's understand the steps first

Key Concepts: Let's go over the following bullet points before we calculate Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors have following components:

- The eigenvector is an array with n entries where n is the number of rows (or columns) of a square matrix. The eigenvector is represented as x.
- Key Note: The direction of an eigenvector does not change when a linear transformation is applied to it.
- Therefore, Eigenvector should be a non-null vector
- We are required to find a number of values, known as eigenvalues such that

$$A x = \lambda x$$

The above equation states that we need to find eigenvalue (λ) and eigenvector (x) such that when we multiply a scalar λ (eigenvalue) to the vector x (eigenvector) then it should equal to the linear transformation of the matrix A once it is scaled by vector x (eigenvector).

Eigenvalues are represented as λ .

Keynote: The above equation should not be invertible.

There are two special keywords which we need to understand: Determinant of a matrix and an identity matrix

- The determinant of a matrix is a number that is computed from a square matrix. In a nutshell, the diagonal elements are multiplied by each other and then they are subtracted together. We need to ensure that the determinant of the matrix is 0.
- Last component: We need an Identity Matrix. An identity square matrix is a matrix that has 1 in diagonal and all of its elements are 0. The identity matrix is represented as I:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

We can represent

$$A * x = \lambda * x$$

As:

$$A * x - \lambda * x = 0$$

Now, we need to compute a characteristic equation.

$$|A - \lambda * I| = 0$$

Subsequently

$$\text{Determinant}(A - \text{Lambda} * I) = 0$$

4.2 How do I calculate Eigenvalue?


The task is to find Eigenvalues of size n for a matrix A of size n.

Therefore, the aim is to find: Eigenvector and Eigenvalues of A such that:

$$A * \text{Eigenvector} - \text{Eigenvalue} * \text{EigenVector} = 0$$

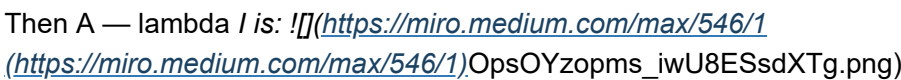
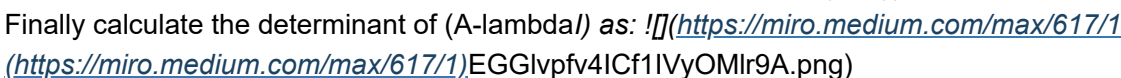
Find Lambda Such that $\text{Determinant}(A - \text{Lambda} * I) = 0$

Based on the concepts learned above:

1. lambda I is: 

If A is:

$$\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}$$

1. Then $A - \text{lambda} I$ is: 
2. Finally calculate the determinant of $(A - \text{lambda} I)$ as: 

Once we solve the equation above, we will get the values of lambda. These values are the Eigenvalues.

I will present a working example below to illustrate the theory so that we understand the concepts thoroughly.

4.3 How do I calculate Eigenvector?

Once we have calculated eigenvalues, we can calculate the Eigenvectors of matrix A by using Gaussian Elimination. Gaussian elimination is about converting the matrix to row echelon form. Finally, it is about solving the linear system by back substitution.

A detailed explanation of Gaussian elimination is out of the scope of this article so that we can concentrate on Eigenvalues and Eigenvectors.

Once we have the Eigenvalues, we can find Eigenvector for each of the Eigenvalues. We can substitute the eigenvalue in the lambda and we will achieve an eigenvector.

$$(A - \text{lambda} * I) * x = 0$$

Therefore if a square matrix has a size n then we will get n eigenvalues and as a result, n eigenvectors will be computed to represent the matrix.

Now that we have the key, it is the time to compute the Eigenvalues and Eigenvectors together.

We will later implement and calculate EigenValues and EigenVectors in Python as we learn python.

Reference and Links to follow:

To learn about Bayesian Statistics, i would suggest to follow this article on AnalyticsVidhya:

<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/>
(<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/>)

Boost your linear algebra skills: <https://towardsdatascience.com/boost-your-data-sciences-skills-learn-linear-algebra-2c30fdd008cf> (<https://towardsdatascience.com/boost-your-data-sciences-skills-learn-linear-algebra-2c30fdd008cf>)

A comprehensive beginners guide to Linear Algebra for Data Scientists -

<https://www.analyticsvidhya.com/blog/2017/05/comprehensive-guide-to-linear-algebra/>
(<https://www.analyticsvidhya.com/blog/2017/05/comprehensive-guide-to-linear-algebra/>)

Video Tutorials to learn Linear Algebra:

Khan Academy has really good tutorials to learn the basics of linear algebra:

<https://www.khanacademy.org/math/linear-algebra> (<https://www.khanacademy.org/math/linear-algebra>)

Personally, i would recommed this playlist by 3Blue1Brown (You probably wouldn't find anyone who explains it better): https://www.youtube.com/watch?v=fNk_zzaMoSs&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab
(https://www.youtube.com/watch?v=fNk_zzaMoSs&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab)