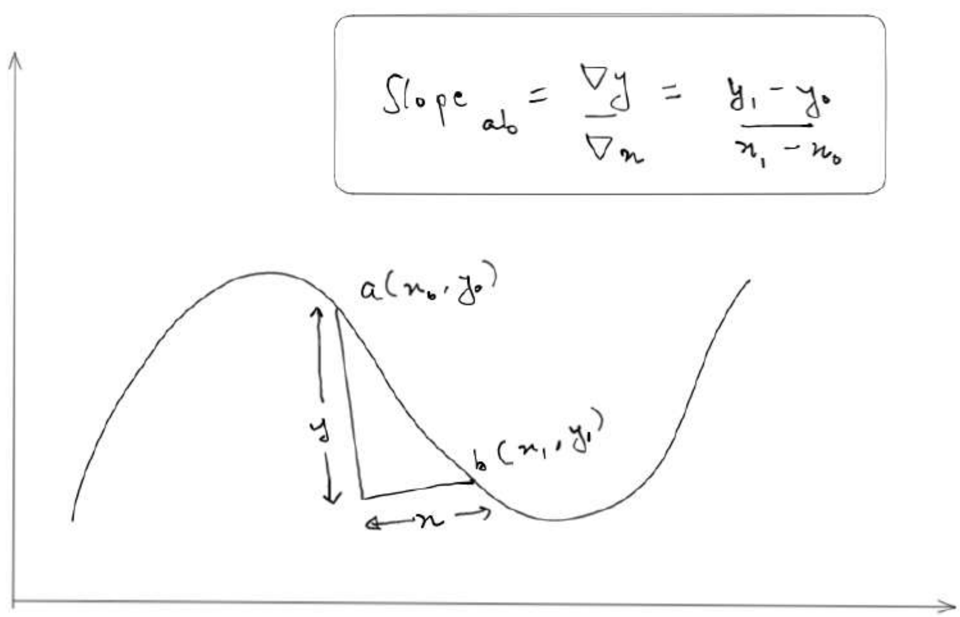
# Calculus Essentials for Machine Learning: Simplifying Complex Models

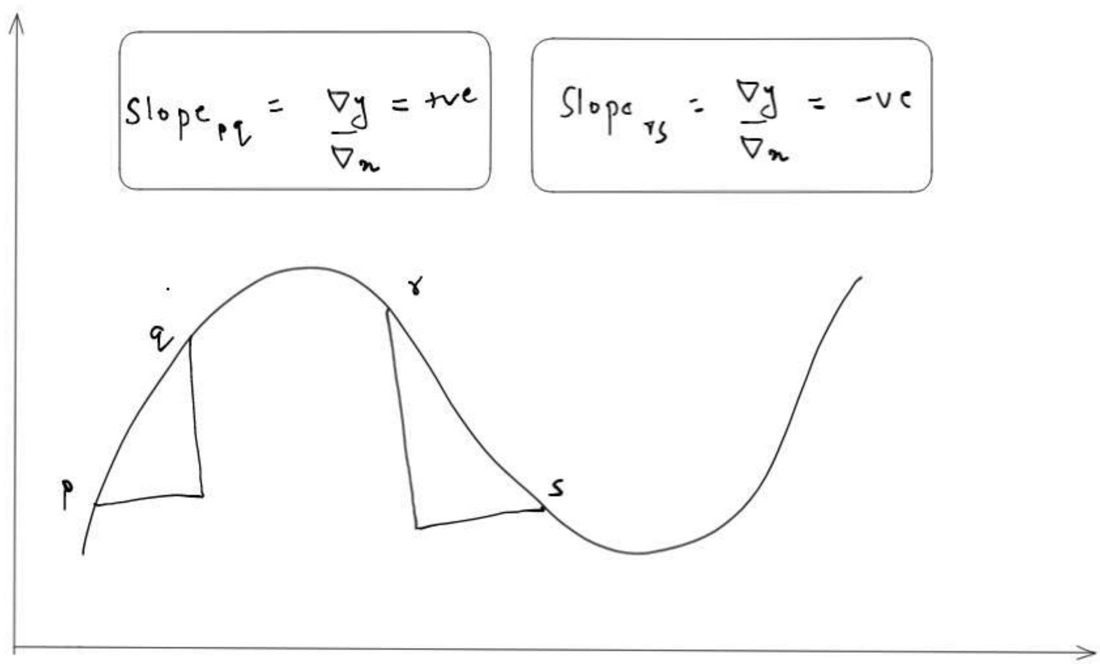
The rate of change of a function with respect to a variable is called differentiation. It finds its applications in various fields, machine learning being one of them. It can be used to find the slope (rate of change) of a function at a specific point in mathematics. However, it is not limited to this; it can also be used to determine the function’s behavior.




Slope of the function between points a&b

Ishan Modi

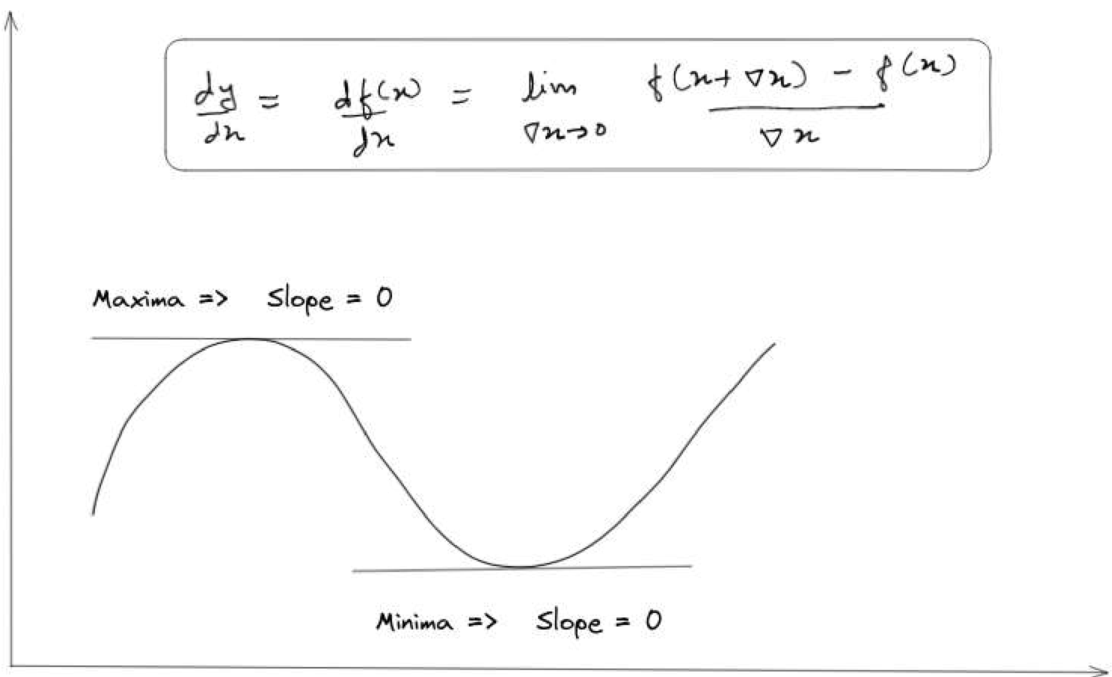
The slope of a function can be negative or positive depending on how it is calculated. The following figure illustrates how the slope changes depending on where we are in the curve. Moving upward in the curve leads to a positive difference between the y-coordinates, which results into a positive slope, and moving downward in the curve leads to in a negative difference between the y-coordinates, which results into a negative slope. Keep in mind that the difference of the x-coordinates remains positive.




positive and negative slope

Finding the minima and maxima of a function and understanding how it behaves are both made possible by the slope of that function. When a function's slope is equal to zero, it reaches its extreme. In other words, the maximum and minimum of a function are the values of the variables (x&y in this case) at which the slope is zero.

The y-coordinate is a function of the x-coordinate in all the mentioned examples. In the figure below, we calculate the slope at a point by taking x-coordinates that are very close to one another. The outcome thus obtained is known as the function's derivative.




maxima and minima and slope of tangent

We have learned up to this point that maxima and minima are obtained when we differentiate the function and obtain slope of tangent (slope at a point). We are still unsure of whether the variables discovered after the first derivative when plugged in the function give us maxima or minima. Double derivative is the answer. When we populate the found variables in the double derivative and the result turns out to be negative, the function attains its maxima and if the result turns out to be positive, the function attains a minima.

Some basic formulas and rules for differentiation are as follows:

$$\frac{d}{dx}\left(cf(x)\right) = cf'(x), \ c \text{ is any constant.} \qquad \left(f(x) \pm g(x)\right)' = f'(x) \pm g'(x)$$

$$\frac{d}{dx}\left(x^n\right) = nx^{n-1}, \ n \text{ is any number.} \qquad \frac{d}{dx}(c) = 0, \ c \text{ is any constant.}$$

$$(f\,g)' = f'g + f\,g' \quad \textbf{– (Product Rule)} \qquad \left(\frac{f}{g}\right)' = \frac{f'g - f\,g'}{g^2} \quad \textbf{– (Quotient Rule)}$$

$$\frac{d}{dx}\left(f\left(g(x)\right)\right) = f'\left(g(x)\right)g'(x) \quad \textbf{(Chain Rule)}$$

$$\frac{d}{dx}\left(\mathbf{e}^{g(x)}\right) = g'(x)\mathbf{e}^{g(x)} \qquad \frac{d}{dx}\left(\ln g(x)\right) = \frac{g'(x)}{g(x)}$$

Derivative cheat sheet

Partial Derivatives is also a widely used concept when it comes to calculus. The core concept of partial derivative revolves around functions having two independent variables. Such functions are partially differentiated with respect to each of the variables taking the other variables as constant.

Ishan Modi

$$P = xy \Rightarrow \frac{dP}{dx} = y + x\frac{dy}{dx}, \quad \frac{dP}{dy} = x + y\frac{dx}{dy} \quad \Big\} \text{Normal Derivative}$$

$$\Rightarrow \frac{\partial P}{\partial x} = y, \quad \frac{\partial P}{\partial y} = x \quad \Big\} \text{Partial Derivative}$$

comparison between normal derivatives and partial derivatives

I hope that this blog has helped you gain a better understanding of calculus and its importance in the field of machine learning. By mastering the principles of calculus, you can take your skills and knowledge to the next level and create innovative solutions that can shape the future of technology. Keep learning, keep exploring, and never stop seeking knowledge!

Ishan Modi

# Unlocking the Power of Linear Algebra: A Journey through Vectors, Matrices, and Beyond

The study of vectors, matrices, and linear transformations is covered in the area of mathematics known as linear algebra. It includes solving systems of linear equations, manipulating and transforming data.

Let us now go over all of the elements that comprise linear algebra, from the most basic to the most complex.

- **Scalar** — single numbers
- **Vector** — a one-dimensional array of numbers
- **Matrix** — a two-dimensional array of numbers
- **Tensor** — more than two-dimensional array of numbers

Now there are different types of operations we can perform on the aforementioned elements.

- **Normal Addition**
  — Done between two or more matrices/vectors.

Ishan Modi

— The sizes of the matrices/vectors that are to be added should be the same.

Addition (both A&B are matrix)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$A+B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} \Rightarrow A_{ij} + B_{ij} = C_{ij}$$

- **Broadcasting**

  — Done between a matrix and a vector.

  — The no of the rows in the matrix should be equal to the number of elements in the vector.

Broadcasting (A is a matrix and b is a vector)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad b = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix}$$

$$A+b = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{11} \\ a_{21} + b_{21} & a_{22} + b_{21} \end{bmatrix} \Rightarrow A_{ij} + b_{j} = C_{ij}$$

- **Normal Multiplication**

  — Done between two or more matrices.

— The number of the columns in the first matrix should be equal to the number of rows in the second matrix.

**Multiplication** (both A&B are matrix)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}_{2\times3} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}_{3\times2}$$

$$A*B = \begin{bmatrix} \vec{a_{1j}} \cdot \vec{b_{i1}} & \vec{a_{1j}} \cdot \vec{b_{i2}} \\ \vec{a_{2j}} \cdot \vec{b_{i1}} & \vec{a_{2j}} \cdot \vec{b_{i2}} \end{bmatrix} \implies A_{ij} * B_{jk} = C_{ik}$$

- **Hadamard Product**

  — Done between two or more matrices. Element-wise multiplication of matrices.

  — The size of the matrices should be the same.

**Hadamard Product** (both A&B are matrix)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$A*B = \begin{bmatrix} a_{11} * b_{11} & a_{12} * b_{12} \\ a_{21} * b_{21} & a_{22} * b_{22} \end{bmatrix} \implies A_{ij} * B_{ij} = C_{ij}$$

Ishan Modi

- **Transpose**
  — Done for one matrix.

Transpose (both A&B are matrix)

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

- **Inverse**
  — Done for one matrix.

Inverse (both A&B are matrix)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Rightarrow A^{-1} = \frac{\text{Adjoint } (A)}{\text{Det } (A)} = \frac{\begin{bmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix}^T}{|A|}$$

Ishan Modi

Till now we have discussed the elements present in linear algebra, let us now discuss how to calculate the magnitude and direction of these elements.

- **Norm** — Used to find the magnitude of vector/matrix.

## Types of Norms

1 - Norm $\Rightarrow$ $$\|X\|_1 = \left(|x_1| + |x_2| + |x_3| + \dots + |x_n|\right)$$

Euclidean Norm $\Rightarrow$ $$\|X\|_2 = \left(x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2\right)^{1/2}$$

2 - Norm

p - Norm $\Rightarrow$ $$\|X\|_p = \left(x_1^p + x_2^p + x_3^p + \dots + x_n^p\right)^{1/p}$$

$\infty$ - Norm $\Rightarrow$ $$\|X\|_\infty = \max\left(|x_1|, |x_2|, |x_3|, \dots, |x_n|\right)$$

Frobenius Norm $\Rightarrow$ $$A_F = \left(\sum_{i,j} a_{ij}^2\right)^{1/2}$$ Used in Matrices

- Dot product — Used to find projection of one vector onto another. Dot product is a scalar produced by vectors of

Ishan Modi

the same dimension. It can also be represented using trigonometric functions.

Dot Product Formula

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Trigonometric Representation

$$\frac{\vec{A} \cdot \vec{B}}{\|A\| \, \|B\|} = \cos\theta$$

- Unit vector — Used to denote the direction of a vector. Magnitude of a unit vector is always one.

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{x_1}{\|x\|} & \frac{x_2}{\|x\|} & \frac{x_3}{\|x\|} \end{bmatrix}$$ Unit Vector

$$\Rightarrow \sqrt{\frac{x_1^2}{\|x\|^2} + \frac{x_2^2}{\|x\|^2} + \frac{x_3^2}{\|x\|^2}}$$ Finding Magnitude of Unit vector

$$\Rightarrow \sqrt{\frac{x_1^2}{x_1^2 + x_2^2 + x_3^2} + \frac{x_2^2}{x_1^2 + x_2^2 + x_3^2} + \frac{x_3^2}{x_1^2 + x_2^2 + x_3^2}}$$ $\|x\|$ is Euclidean Norm

$$\Rightarrow 1$$

Another important topic in linear algebra is Linear independence of vectors. A set of vectors is said to be linearly independent if none of the vectors in the set can be expressed as a linear combination of the others. In other words, a set of vectors is linearly independent if the only way to get the zero vector as a linear combination of those vectors is by setting all the coefficients to zero.

With this we come to an end of this blog. I hope that this blog has given you a solid foundation to build upon and inspired you to explore various applications of linear algebra.

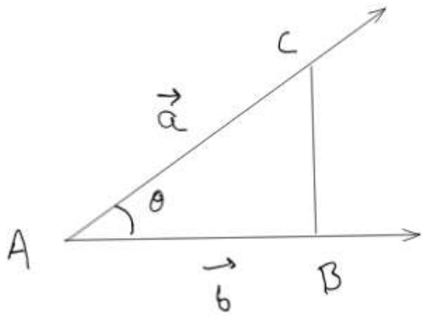
Ishan Modi

# Mastering the Fundamentals: Trigonometry, Logarithms and Exponents in Machine Learning

Trigonometry is a branch of mathematics that deals with the study of relationships between angles and sides of triangles. It is a fundamental topic in mathematics and has applications in various fields, including Machine learning. A lot of vector calculus and linear algebra are used in this field of study and the relationship between these vectors is often derived using trigonometry.

Three fundamental functions — sine, cosine, and tangent — must be understood in order to comprehend trigonometry. The following are the formulas for the aforementioned functions.

Ishan Modi




$$\text{Sine} = \frac{\text{Opposite}}{\text{Hypotenuse}} = \frac{BC}{AC}$$

$$\text{Cosine} = \frac{\text{Adjacent}}{\text{Hypotenuse}} = \frac{AB}{AC}$$

$$\text{Tangent} = \frac{\text{Opposite}}{\text{Adjacent}} = \frac{BC}{AB}$$

Trigonometric functions

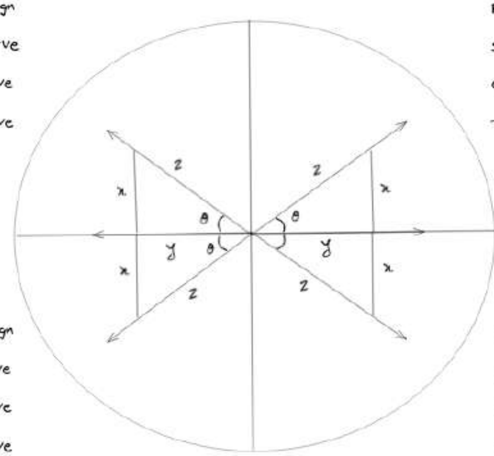
When these functions are applied to angles, a specific value is produced. These numbers change as the angle changes. These functions have a special relationship with the quadrants of the Cartesian Coordinate System.

Quadrant 2

| Function | x | y | z | Formula | Sign |
|---|---|---|---|---|---|
| Sine | +ve | +ve | +ve | x/z | +ve |
| Cosine | +ve | -ve | +ve | y/z | -ve |
| Tangent | +ve | -ve | +ve | x/y | -ve |

Quadrant 1

| Function | x | y | z | Formula | Sign |
|---|---|---|---|---|---|
| Sine | +ve | +ve | +ve | x/z | +ve |
| Cosine | +ve | +ve | +ve | y/z | +ve |
| Tangent | +ve | +ve | +ve | x/y | +ve |




Quadrant 3

| Function | x | y | z | Formula | Sign |
|---|---|---|---|---|---|
| Sine | -ve | +ve | +ve | x/z | -ve |
| Cosine | +ve | -ve | +ve | y/z | -ve |
| Tangent | -ve | -ve | +ve | x/y | +ve |

Quadrant 4

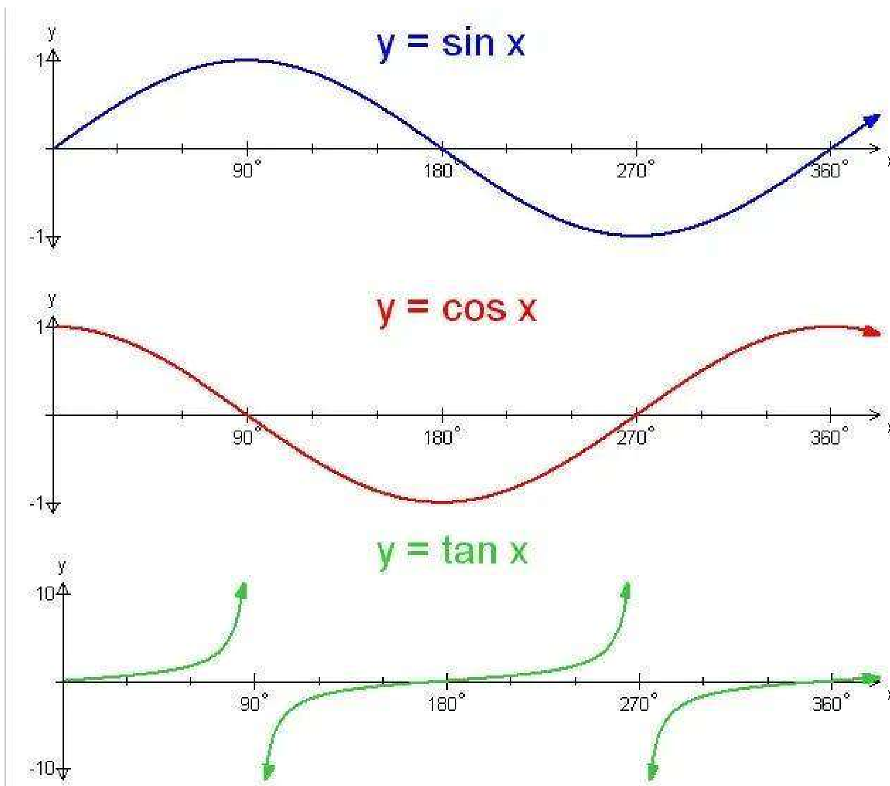
| Function | x | y | z | Formula | Sign |
|---|---|---|---|---|---|
| Sine | -ve | +ve | +ve | x/z | -ve |
| Cosine | +ve | +ve | +ve | y/z | +ve |
| Tangent | -ve | +ve | +ve | x/y | -ve |

Ishan Modi

Applying the logic outlined above, we can arrive at the relationship shown below

- **Sine** — positive in 1st and 2nd quadrant and negative in 3rd and 4th quadrant
- **Cosine** — positive in 1st and 4th quadrant and negative in 2nd and 3rd quadrant.
- **Tangent** — positive in 1st and 3rd quadrant and negative in 2nd and 4th quadrant.

The two-dimensional graph showing the pictorial representation of the previously discussed trigonometric functions is shown below.




Logarithms are a mathematical concept that represent the inverse operation of exponentiation. They have many applications in various fields, including machine learning, engineering, finance, and computing.

Logarithms are used to simplify complex calculations, especially when dealing with large numbers or when multiplying or dividing numbers with different powers. By converting multiplication and division into addition and subtraction, logarithms can make complex calculations more manageable. They are also used to analyze algorithms, optimize data structures, and to measure the complexity of algorithms.

## Common logarithmic Formulas

$$\log_a x = y \Rightarrow x = a^y$$

$$\log_a b = \frac{\ln a}{\ln b} \text{ or } \frac{\log_{10} a}{\log_{10} b}$$

### Natural Logarithm

$$\ln xy = \ln x + \ln y$$
$$\ln \frac{x}{y} = \ln x - \ln y$$
$$\ln x^y = y \cdot \ln x$$
$$\ln e^x = x$$
$$e^{\ln x} = x$$

### Logarithm with base a

$$\log_a xy = \log_a x + \log_a y$$
$$\log_a \frac{x}{y} = \log_a x - \log_a y$$
$$\log_a x^y = y \cdot \log_a x$$
$$\log_a a^x = x$$
$$a^{\log_a x} = x$$

Ishan Modi

Exponents, or powers, are a fundamental mathematical concept that plays an important role in various aspects of machine learning, including feature scaling, regularization, loss functions, and neural networks.

$$a^m a^n = a^{m+n}$$

$$\frac{a^m}{a^n} = a^{m-n},\ a \neq 0$$

$$a^{\frac{1}{n}} = \sqrt[n]{a}$$

$$(ab)^m = a^m b^m$$

$$a^{-m} = \frac{1}{a^m},\ a \neq 0$$

$$a^{\frac{m}{n}} = \sqrt[n]{a^m} = \left(\sqrt[n]{a}\right)^m$$

$$(a^m)^n = a^{mn}$$

$$\left(\frac{a}{b}\right)^m = \frac{a^m}{b^m},\ b \neq 0$$

$$a^0 = 1,\ a \neq 0$$

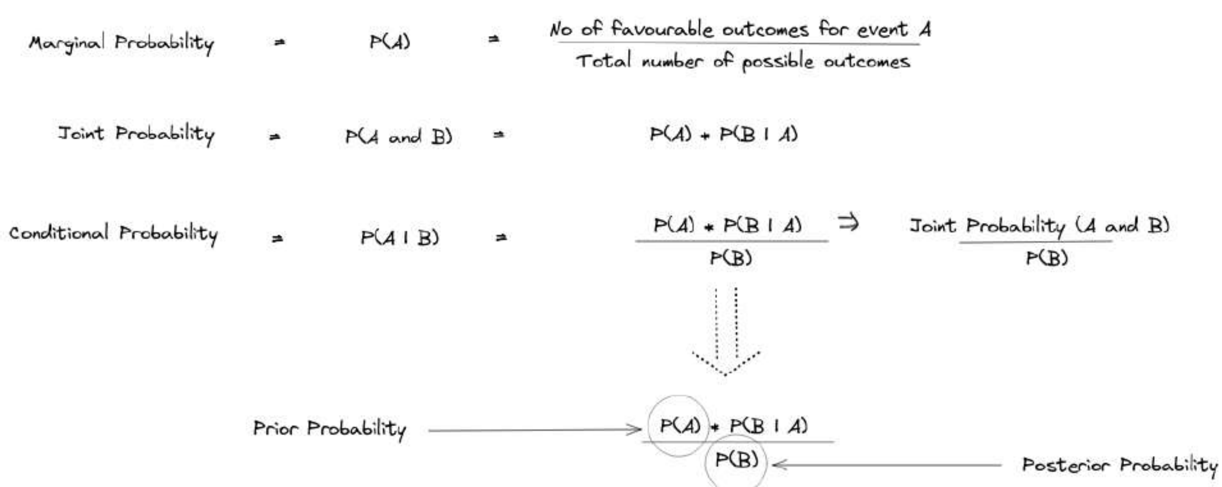
Exponent Formulas

That's All, Happy Learning !!!

Ishan Modi

# Probability and Statistics essentials for Machine Learning Practitioners

Probability forms the basis of almost all machine learning algorithms. It finds its applications in classical machine learning algorithms like Bayesian theorem to deep learning in the form of activation functions. The probability of an event is calculated by dividing the number of favorable outcomes by the total number of possible outcomes. This is known as the classical definition of probability. However, in real-world scenarios, probability can be more complex.




Types of Probabilities

Ishan Modi

Statistics is a branch of mathematics that deals with collecting, analyzing, interpreting, and presenting data. It involves the use of mathematical tools and techniques to make sense of data and to draw conclusions about populations based on samples. It is an essential component of machine learning used in providing a foundation for data analysis, model selection, inference, and interpretation.

Statistics is often used to measure the central tendency of the data. This can be done using the following concepts.

- **Mean** — The arithmetic average of the dataset.
- **Median** — The central value of the dataset.
- **Mode** — The most frequently occurring value in the dataset.

Mean

$$\boxed{\frac{\text{Sum of all elements in the dataset}}{\text{Total number of elements in the dataset}}}$$

Example

Data = 5, 10, 20, 30, 35

$$\text{Mean} = \frac{5+10+20+30+35}{5}$$

= 20

Median

Median = $\left(\frac{\text{Total Elements} + 1}{2}\right)^{TH}$ Element — If size of dataset is ODD

= AVG $\left(\left(\frac{\text{Total Elements}}{2}\right)^{TH}\text{Element}, \left(\frac{\text{Total Elements}}{2} + 1\right)^{TH}\text{Element}\right)$ — If size of dataset is EVEN

Example

Data = 5, 10, 20, 30, 35

Median = $\left(\frac{5+1}{2}\right)^{th}$ element

= 20

Data = 5, 10, 20, 30, 35, 40

Median = Avg $\left(\left(\frac{6}{2}\right)^{th}\text{element}, \left(\frac{6}{2}+1\right)^{th}\text{element}\right)$

= 25

Mode

Example

Data = 5, 5, 5, 30, 35

Mode = 5

Ishan Modi

CloudTech

Now we move on to leveraging statistical methods to derive relation in the data. This can be done using the following concepts

- **Variance** —Variance is a statistical measure that quantifies the spread or variability of a set of data points around its mean or expected value.
  A high variance indicates that the data points are spread out over a wider range of values, while a low variance indicates that the data points are clustered more tightly around the mean.
- **Standard Deviation** —Standard deviation is a statistical measure that describes the amount of variation or dispersion of a set of data points from its mean or expected value.
  A high standard deviation indicates that the data points are more spread out from the mean, while a low standard deviation indicates that the data points are closer to the mean.
- **Covariance** —Covariance is a statistical measure that quantifies the degree to which two random variables in a dataset vary together.
  A positive covariance indicates that the two variables tend to move in the same direction, while a negative covariance indicates that they tend to move in opposite directions. A covariance of zero indicates that the variables are not related.

Ishan Modi

- **Correlation** —Correlation is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous random variables. A positive correlation indicates that as one variable increases, the other variable tends to increase as well, while a negative correlation indicates that as one variable increases, the other variable tends to decrease. A correlation of zero indicates that the two variables are not related. The range of correlation is between -1 to 1.

$$\text{Variance } (x) = \frac{(x_i - \bar{x})^2}{N}$$

$$\text{Standard Deviation } (x) = \sqrt{\frac{(x_i - \bar{x})^2}{N}} = \sqrt{\text{Variance}}$$

$$\text{Covariance } (x \,\&\, y) = \frac{(x_i - \bar{x})(y_i - \bar{y})}{N} \quad \text{(x and y are attributes/columns of the dataset)}$$

$$\text{Correlation } (x \,\&\, y) = \frac{\text{Covariance } (x \,\&\, y)}{\sqrt{\text{Variance}(x) * \text{variance}(y)}} \quad \text{(x and y are attributes/columns of the dataset)}$$

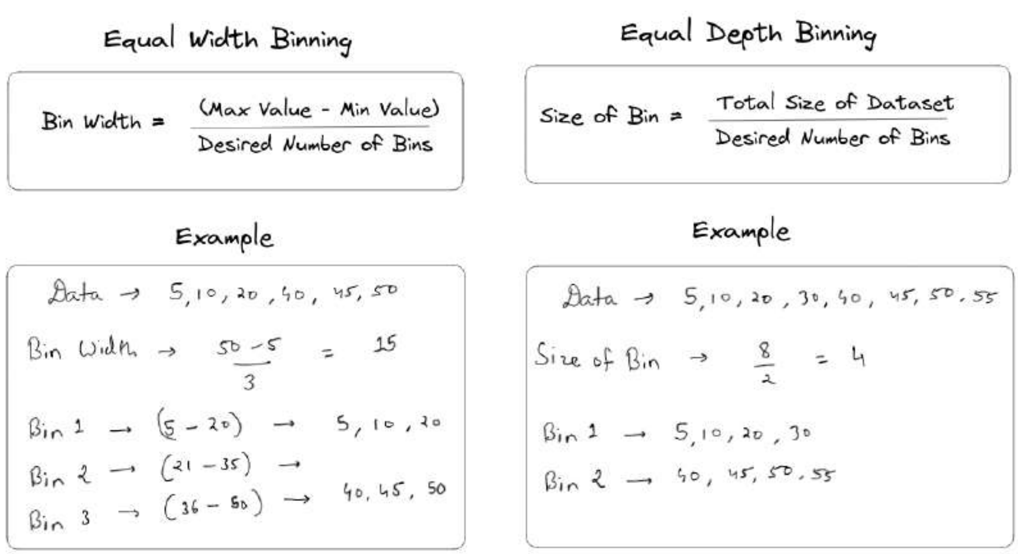
With this we conclude this blog on Probability and Statistical methods for better data interpretation. Hope you found it useful !!!

Ishan Modi

# From Messy to Meaningful: The Math behind Data Cleaning and Normalization in Data Analysis

Binning in Exploratory Data Analysis (EDA) is a process of grouping a continuous variable into a discrete number of categories or bins. It is a useful tool in data analysis that can help in simplifying, organizing, and interpreting complex data sets. However, it is important to choose the appropriate binning method and the number of bins carefully to avoid loss of information or misleading results.

Following are the commonly used binning techniques




Ishan Modi

Smoothing in Exploratory Data Analysis (EDA) is a process of removing noise or fluctuations from a data set to reveal a clearer underlying trend or pattern. The main purpose of smoothing is to simplify the data and reduce the effect of small variations in the data. It involves applying mathematical functions to the data to create a smoothed curve or line that approximates the underlying trend. It is generally used after the process of Binning.

Following are the commonly used smoothing techniques

**Mean Smoothing**

$$\text{Mean} = \frac{\text{Sum of elements in Bin}}{\text{Total Size of Bin}}$$

**Example**

Bin 1 → 5, 10, 20, 30

Bin 2 → 40, 45, 50, 55

Mean Bin 1 → Avg (5, 10, 20, 30)

Mean Bin 2 → Avg (40, 45, 50, 55)

Bin 1 → 16.2, 16.2, 16.2, 16.2

Bin 2 → 47.5, 47.5, 47.5, 47.5

**Median Smoothing**

$$\text{Median} = \left(\frac{\text{Total Elements} + 1}{2}\right)^{TH} \text{Element}$$

If size of dataset is ODD

$$= \text{AVG}\left(\left(\frac{\text{Total Elements}}{2}\right)^{TH} \text{Element}, \left(\frac{\text{Total Elements}}{2} + 1\right)^{TH} \text{Element}\right)$$

If size of dataset is EVEN

**Example**

Bin 1 → 5, 10, 20, 30

Bin 2 → 40, 45, 50, 55

Median Bin 1 → $\frac{10 + 20}{2} = 15$

Median Bin 2 → $\frac{45 + 50}{2} = 47.5$

Bin 1 → 15, 15, 15, 15

Bin 2 → 47.5, 47.5, 47.5, 47.5

Ishan Modi

When it comes to deep learning, normalization is a critical component. As the name “normalization” implies, this process involves converting arbitrary data to a specified scale. In plainer terms, it is employed to cap the data. Normalization makes data more understandable while preserving the it’s original information.

Following are a couple of advantages of using normalization in deep learning

- The data used at every step in one’s model comes from the same distribution, this makes it easier for the model to develop relations and thus, learn better.
- The complexity of the calculations is reduced because normalization scales the number to the ideal amount where the values in the data are neither too small nor too large.

Ishan Modi

Min-Max Scaling $\Rightarrow$

$$x_i' = \left[\frac{x_i - Min(x)}{Max(x) - Min(x)} \times (new_{max} - new_{min})\right] + new_{min}$$

Min and Max value of the scaled data

Z-Score Normalization $\Rightarrow$

$$x_i' = \frac{x_i - \bar{x}}{\sigma_x}, \quad \text{where } \sigma_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}}$$

$$x_i' = \frac{x_i - \bar{x}}{S_x}, \quad \text{where } S_x = \frac{1}{N}\left(\Sigma|x_i - \bar{x}|\right)$$

Decimal Scaling $\Rightarrow$

$$x_i' = \frac{x_i}{10^j}, \quad \text{where } j = \text{maximum number of digits in a given data}$$

Types of normalization

Apart from the aforementioned data-cleaning and normalization techniques, there are numerous other things that you should be aware of if you work in a professional setting. Handling issues with missing data, duplicate records, inconsistent data, mismatched data types, and much more.

That’s all from this blog, Until next time !!!

Ishan Modi