Norms

Vector Norms are defined as a set of functions that take a vector as an input and output a **positive value** against it. This is called the *magnitude* of a vector. We can obtain different lengths for the same vector depending on the type of function we use to calculate the magnitude.

Backpropagation is a gradient estimation method used in training neural networks.

It computes the gradient of a loss function with respect to the weights of the network.

The goal is to adjust the network's parameters (weights) to minimize the loss function, improving the model's predictions.

Essentially
is the avera
This scale
below:

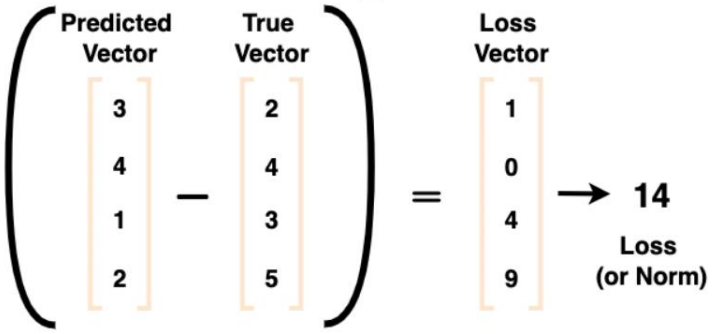


Fig: A diagram showing the calculation of loss using predicted values and true values

The Standard Norm Equation — P-norm

All norm functions originate from a standard equation of Norm, known as the **p-norm**. For different values of the parameter p (p should be a real number greater than or equal to 1), we obtain a different norm function. The generalized equation, however, is shown below:

The p-norm equation:

This takes an n-dimensional vector x and raises each element to its p-th power. Then, we sum all the obtained elements and take the p-th root to get the p-norm of the vector, also known as its magnitude. Now, with different values of the parameter p, we will obtain a different norm function. Let's discuss them one by one below.

$$\underbrace{\|x\|_p}_{p-Norm} = \left(\sum_i^n |x_i|^p\right)^{\frac{1}{p}} = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$$

L0 Norm:

Although p=0 lies outside the domain of the p-norm function, substituting p=0 in the above equation gives

us the individual vector elements raised to the power 0, which is 1 (provided the number is not zero). Furthermore, we also have a p-th root in the equation, which is not defined for p=0. To handle this, the standard way of defining the L0 norm is to count the number of non-zero elements in the given vector. The image below shows the output of the L-0 norm function for the given vector:

L1 Norm:

Substituting p=1 in the standard equation of p-norm, we get the following:

$$\underbrace{\|x\|_1}_{L1\ Norm} = (\sum_i^n |x_i|) = (|x_1| + |x_2| + \cdots + |x_n|)$$

- When used to compute the loss, the L1 norm is also referred to as the Mean Absolute Error.
- L1 norm varies linearly for all locations, whether far or near the origin.

The image below shows the output of the L1 norm function for the given vector:

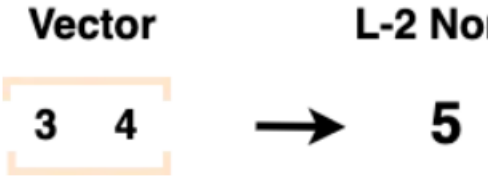
L2 Norm:

Of all norm functions, the most common and important is the L2 Norm. Substituting p=2 in the standard equation of p-norm, which we discussed above, we get the following equation for the L2 Norm:

$$\underbrace{\|x\|_2}_{L-2\ Norm} = (\sum_i^n |x_i|^2)^{\frac{1}{2}} = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

- The above equation is often referred to as the root mean squared error when used to compute the error.
- L2 norm measures the distance from the origin, also known as Euclidean distance.

The image below shows the output of the L2 norm function for the given vector:



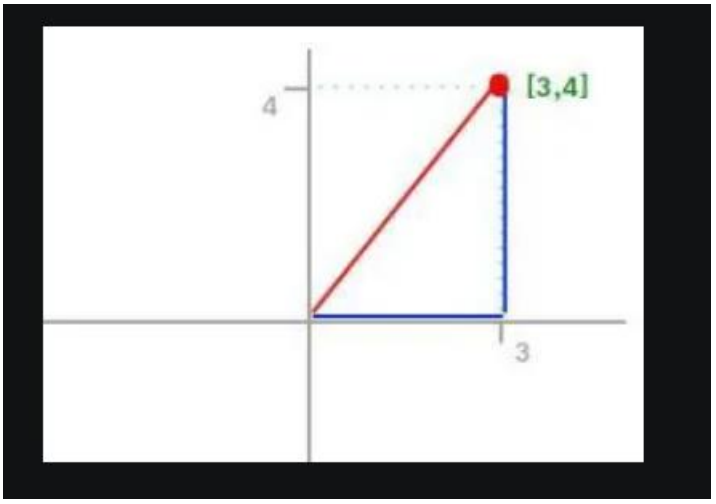


Fig: Red line is Euclidean distance

Squared L2 Norm:

As the name indicates, the squared L2 Norm is the same as the L2 Norm but squared.

$$\underbrace{\|x\|_2^2}_{Squared\ L2\ Norm} = \left(\sum_i^n |x_i|^2\right) = \left(|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2\right)$$

- The above equation is often referred to as the mean squared error when used to compute the error in machine learning.
- The squared L2 Norm is relatively computationally inexpensive to use compared to the L2 Norm. This is because:
    - It is missing the square root.
    - Within Machine Learning applications, the derivative of the Squared L2 Norm is easier to compute and store. The derivate of an element in the Squared L2 Norm requires the element

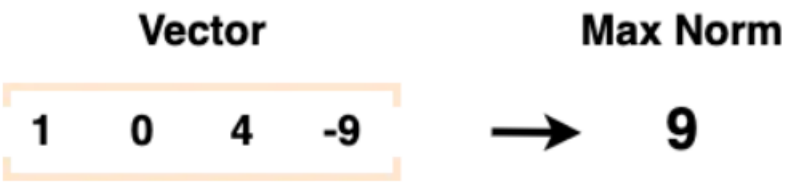
itself. However, in the case of the L2 Norm, the entire vector is needed.

Max Norm (or L-∞ Norm):

As infinity is an abstract concept in Mathematics, we can't just substitute p=∞ in the standard p-norm equation. However, we can study the function's behavior as p approaches infinity using limits. A simple derivation for the equation of Max-norm can be found here.

$$\underbrace{\|x\|_\infty}_{Max\ \ Norm} = \max_i |x_i|$$

- Max norm returns the absolute value of the largest magnitude element.
- The image below shows the output of the Max norm function for the given vector:



Conclusion:

- Vector norm is a function that takes a vector as an input and outputs a positive value.
- All norm functions can be derived from a single equation. The family of norm functions is known as p-norm.
- The L1 norm is also referred to as the Mean Absolute Error.
- The L2 Norm is also referred to as the Root Mean Squared Error.
- The Squared L2 Norm is also referred to as the Mean Squared Error.