

Objectives

- Norms
 - Inner Product
 - Projection
-

1 Norms

A Norm is a function that takes a vector and maps it to a real number. It is closely tied to the concepts of absolute value and length. In relation to distance, $\|x\| = \|x - 0\| = d(x, 0)$.

The Norm function is $\|\cdot\| : V \rightarrow \mathbb{R}^+$ such that:

1. $\|x\| \geq 0$ norms are greater than or equal to 0
2. $\|x\| = 0$ iff $x = 0$ norms are equal to 0 if and only if $x = 0$
3. $\|Cx\| = |C| \cdot \|x\|$ where $C \in \mathbb{R}$
4. $\|x + y\| \leq \|x\| + \|y\|$

Consider the following example.

$$d_H \left(\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right) = 2$$

The Hamming Distance must satisfy all properties in order to be considered a norm function. Consider if x and y were scaled by an arbitrarily chosen constant 3.

$$d_H \left(\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix} \right) = 2$$

The Hamming Distance returns the number of elements that differ between two vectors. The result remains the same despite scaling the vectors. Therefore, the Hamming Distance is **not** a Norm function as it does not meet the third criteria.

$$d_H(Cx, Cy) \neq C \cdot d_H(x, y)$$

There are several types of norms, and we differentiate them by subscripts. The first set of norms are the Minkowski norms or p norms, which are denoted as such:

$$\text{1-norm} = \|x\|_1$$

$$\text{2-norm} = \|x\|_2$$

$$\text{3-norm} = \|x\|_3$$

$$\text{p-norm} = \|x\|_p$$

These norms are calculated differently depending on the subscripts. The equations are shown below. In the first two cases, p is the dimensionality of \mathbf{x} , and in the third dimension, n is the dimensionality of \mathbf{x} .

$$\|x\|_1 = \sum_{i=1}^p |x_i|$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^p x_i^2}$$

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$$

Chebyshev's norm also referred to as the ∞ norm or the max norm is denoted as $\|x\|_\infty$. This norm finds the component of the vector that is the farthest from 0 and returns that. The technical definition is as follows.

$$\|x\|_\infty = \max(|x_1|, |x_2|, |x_3|, \dots, |x_p|)$$

All norms are equivalent but not equal. If $\|x\|_1$ is larger than $\|y\|_1$ then $\|x\|_p$ is larger than $\|y\|_p$.

2 Inner Product

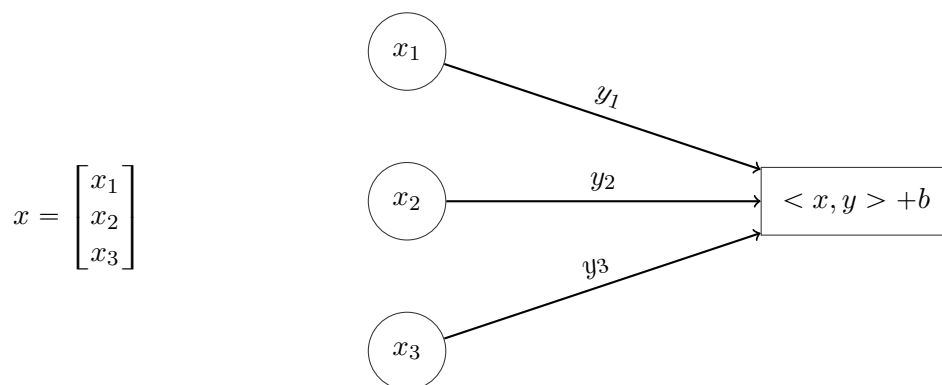
The inner product is $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ such that:

1. $\langle x, x \rangle \geq 0$
2. $\langle x, x \rangle = 0$ iff $x = 0$
3. $\langle x, y \rangle = \langle y, x \rangle$
4. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$, $\alpha, \beta \in \mathbb{R}$, $x, y, z \in V$

This is the basis of the neural networks. A few more aspects of the inner product are:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^P x_i y_i$$

In the diagram below, the vector x contains three input features, and each connection to the output node is weighted by a corresponding component of y . The output is the result of computing the dot product $\langle x, y \rangle$, and then adding a bias b . This is the fundamental operation of a single neuron in a neural network.



An important special case of the inner product occurs when a vector is dotted with itself. This gives us the 2-norm, of the vector:

$$||x|| = \sqrt{\langle x, x \rangle}$$

3 Projection

Projection is used in principle component analysis. We started by looking at the problem to show a situation in which projection might be necessary.

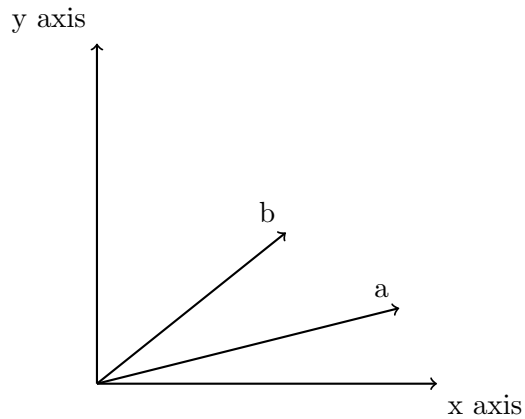
$$Ax = b$$

A is a 2d matrix, b is a known vector, and x is an unknown vector.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix}$$

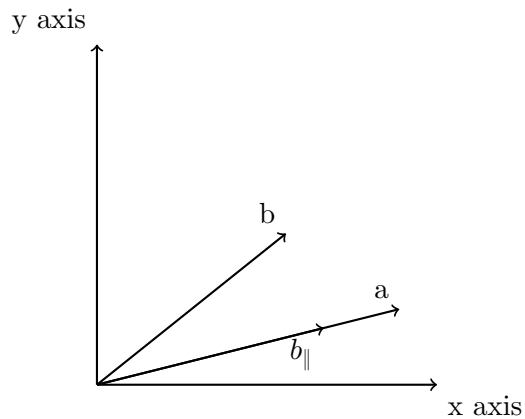
This problem has us trying to find the vector x. This problem may have no perfect solution. This is the sort of problem where we might use projection to solve it because projection is the best solution we can come up with.

He then showed another problem.

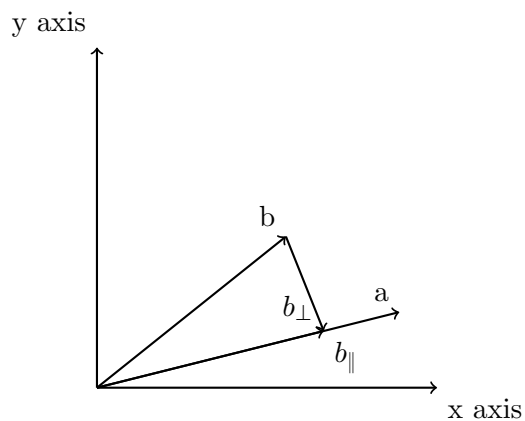


The goal was to project b down onto a . This is necessary because there is no way to turn a into b through multiplication. Multiplication just stretches vector a . There are a few terms we'll need to know before we can do this.

p = projection which is equivalent to b_{\parallel} which is parallel to a . It is also equivalent to a times some constant as it is just a resizing of a or $b_{\parallel} = ca = p$. This is the vector added below.



b_{\perp} is perpendicular to a . It is also known as the residual and is equal to b minus a times some constant c , or $b_{\perp} = b - ca$. This is the vector added below.



It is also important to note that $b = b_{\parallel} + b_{\perp}$ and that $\langle b_{\parallel}, b_{\perp} \rangle = 0$ we use all of these definitions and the inner product to solve it.

$$\begin{aligned}
 0 &= \langle ca, b - ca \rangle \\
 0 &= (ca)^T (b - ca) \\
 0 &= ca^T (b - ca) \\
 0 &= ca^T b - c^2 a^T b \\
 ca^T a &= a^T b \\
 c &= \frac{a^T b}{a^T a} \\
 c &= \frac{\langle a, b \rangle}{\langle a, a \rangle} \\
 ca &= \left(\frac{a^T b}{a^T a} \right) a = \left(\frac{a^T b}{a^T a} \right) a
 \end{aligned}$$

All of this allows you to project b onto a .