Applied Machine Learning for Educational Data Science
Warm-up I Linear Algebra

true

05/26/2021

Contents

[Updated: Wed, May 26, 2021 - 12:14:26]

This is a crash course on Linear Algebra for students in social science. This document does not replace a course on Linear Algebra, and only serves the purpose of developing some vocabulary you will likely encounter in any Machine Learning course or textbook. If you had taken Linear Algebra before, it will refresh your memory.

Vectors and Scalars

A vector is an object with two properties: magnitude and direction. A scalar is an object with magnitude but no direction. Vectors are written in **bold** while scalars are written in *italic* do distinguish the two. Vectors can be shown in *column* format or *row* format.

For instance, $\mathbf{a_1}$ and $\mathbf{a_2}$ below represent row vectors in $\mathbb{R}^2$, z represents a scalar, and $\mathbf{b_1}$ and $\mathbf{b_2}$ represent column vectors in $\mathbb{R}^3$.

$$\mathbf{a_1} = \begin{bmatrix} 3 & 5 \end{bmatrix}$$

$$\mathbf{a_2} = \begin{bmatrix} -2 & 8 \end{bmatrix}$$

$$\mathbf{b_1} = \begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}$$

$$\mathbf{b_2} = \begin{bmatrix} 6 \\ 8 \\ 2 \end{bmatrix}$$

$$z = 5$$

$\mathbb{R}^n$ indicates that the elements of a vector are real numbers and the vector exists in an n-dimensional space. For instance, $\mathbf{a_1}$ and $\mathbf{a_2}$ can be represented in a two-dimensional space while $\mathbf{b_1}$ and $\mathbf{b_2}$ can be represented in a three-dimensional space.
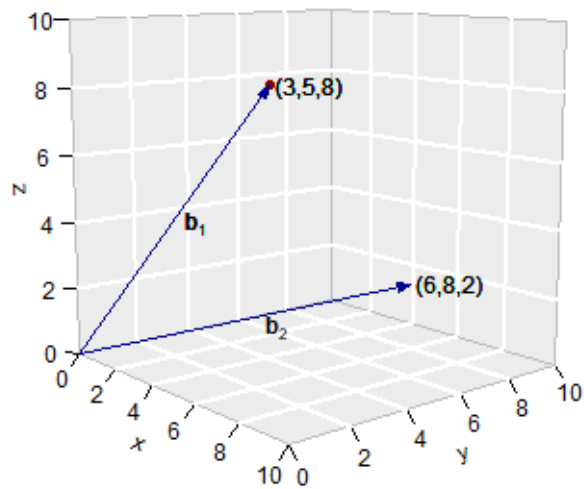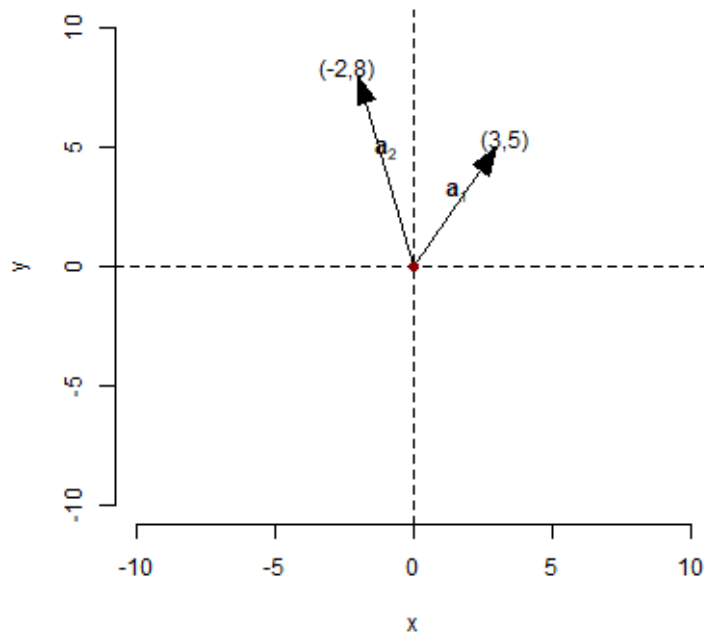
In R, when one creates a vector using the `c()` function, the object operates as a *column vector*. Therefore, when one transpose the object created by the `c()` function, it returns a *row vector*.

$$
\begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}^{T} = \begin{bmatrix} 3 & 5 & 8 \end{bmatrix}
$$

```r
vec1 <- c(3, 5, 8)

vec1    #  this operates as a column vector
```

```
[1] 3 5 8
```

```r
vec2 <- t(vec1)  # transpose the first vector

vec2
```

```
     [,1] [,2] [,3]
[1,]    3    5    8
```

```r
dim(vec2)  # Notice that vec2 has dimensions of 1 x 3.
```

```
[1] 1 3
```

Euclidean Space

In its most general form, an n-dimensional vector of $\mathbf{v}$ in $\mathbb{R}^n$ can be written as,

$$
\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ . \\ . \\ . \\ v_n \end{bmatrix}.
$$

$\mathbb{R}$ indicates that all the elements of this vector $(v_1, v_2, ..., v_n)$ are real numbers, and $\mathbb{R}^n$ represents the set of all n-dimensional vectors. We can also call $\mathbb{R}^n$ a vector space, or more specifically an Euclidean space.

Properties

The following properties hold for vectors that exist in $\mathbb{R}^n$:

1. For all $\mathbf{x} \in \mathbb{R}^n$, there exists an additive identity such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$.

```r
vec1  <- c(3, 5, 8, 10, 9)
zeros <- c(0, 0, 0, 0, 0)

vec1 + zeros
```

```
[1]  3  5  8 10  9
```

2. For all $\mathbf{x} \in \mathbb{R}^n$, there exists an additive inverse such that $\mathbf{x} + -\mathbf{x} = \mathbf{0}$.

```
vec1          <- c(3, 5, 8, 10, 9)
vec1.inverse <- c(-3, -5, -8, -10, -9)

vec1 + vec1.inverse
```

```
[1] 0 0 0 0 0
```

3. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the following commutative law holds:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$$

.

```
vec1          <- c(3, 5, 8, 10, 9)
vec2          <- c(1, 10, 7, 3, 5)

vec1 + vec2
```

```
[1]   4 15 15 13 14
```

```
vec2 + vec1
```

```
[1]   4 15 15 13 14
```

4. For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, the following associative law holds:

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}) = \mathbf{y} + (\mathbf{x} + \mathbf{z})$$

.

```
vec1          <- c(3, 5, 8, 10, 9)
vec2          <- c(1, 10, 7, 3, 5)
vec3          <- c(-2, -4, 7, 3, -5)


vec1 + (vec2 + vec3)
```

```
[1]   2 11 22 16  9
```

```
vec2 + (vec1 + vec3)
```

```
[1]   2 11 22 16  9
```

```
vec3 + (vec1 + vec2)
```

```
[1]   2 11 22 16  9
```

5. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $k \in \mathbb{R}$, the following distributive law holds:

$$k(\mathbf{x} + \mathbf{y}) = k\mathbf{x} + k\mathbf{y}$$

.

```
vec1           <- c(3, 5, 8, 10, 9)
vec2           <- c(1, 10, 7, 3, 5)
k              <- 5

k*(vec1 + vec2)
```

```
[1] 20 75 75 65 70
```

```
k*vec1 + k*vec2
```

```
[1] 20 75 75 65 70
```

6. For all $\mathbf{x} \in \mathbb{R}^n$ and $k, u \in \mathbb{R}$, the following distributive law also holds:

$$(k + u)\mathbf{x} = k\mathbf{x} + u\mathbf{x}$$

.

```
vec1           <- c(3, 5, 8, 10, 9)
k              <- 5
u              <- 10

(k+u)*(vec1)
```

```
[1]   45   75 120 150 135
```

```
k*vec1 + u*vec1
```

```
[1]   45   75 120 150 135
```

7. For all $\mathbf{x} \in \mathbb{R}^n$ and $k, u \in \mathbb{R}$, the following associative law holds:

$$(ku)\mathbf{x} = k(u\mathbf{x}) = u(k\mathbf{x})$$

.

```
vec1           <- c(3, 5, 8, 10, 9)
k              <- 5
u              <- 10

(k*u)*(vec1)
```

```
[1] 150 250 400 500 450
```

```
k*(u*vec1)
```

```
[1] 150 250 400 500 450
```

```
u*(k*vec1)
```

```
[1] 150 250 400 500 450
```

8. For all $\mathbf{x} \in \mathbb{R}^n$, there is a neutral element,

$$\mathbf{1x} = \mathbf{x}$$

.

Inner (Dot) Product

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and defined as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}.$$

The dot product of $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + ... + x_n y_n.$$

```
vec1         <- c(3, 5, 8, 10, 9)
vec2         <- c(1, 10, 7, 3, 5)

sum(vec1*vec2)
```

```
[1] 184
```

Note that both vectors must have the same dimensions, and the inner product of two vectors is always a scalar regardless of their dimensions. You can also use **%*%** to obtain the inner product of two vectors.

```
vec1%*%vec2
```

```
      [,1]
[1,]   184
```

Orthogonal vectors

If the inner product of two vectors is equal to 0, then the vectors are said to be orthogonal (or perpendicular). For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if the inner product is zero,

$$\mathbf{x} \cdot \mathbf{y} = 0.$$

Notice that the following two vectors are orthogonal to each other.

```
vec1          <- c(4,6)
vec2          <- c(-6,4)

vec1%*%vec2
```

```
     [,1]
[1,]    0
```
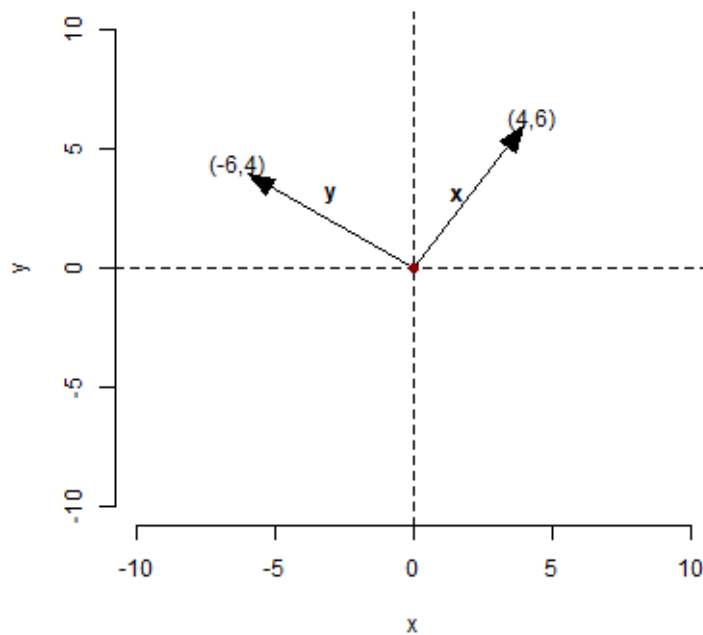

The Euclidean norm of a vector

For all $\mathbf{x} \in \mathbb{R}^n$, the Euclidean norm of a vector is denoted as $\|\mathbf{x}\|$ and calculated as,

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}.$$

The Euclidean norm of a vector is also known as the length of a vector.

```
vec1          <- c(3, 5, 8, 10, 9)
sqrt(sum(vec1^2))
```

```
[1] 16.70329
```

Note that the Euclidean norm of a vector can also be represented as the dot product of a vector by itself,

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

```
vec1         <- c(3, 5, 8, 10, 9)
sqrt(vec1%*%vec1)
```

```
        [,1]
[1,] 16.70329
```

The Euclidean norm (length) of a vector will be always a positive number by definition.

The distance between two vectors

Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in $\mathbb{R}^n$. The distance between $\mathbf{x}$ and $\mathbf{y}$ is denoted as $d(\mathbf{x}, \mathbf{y})$ and calculated as,

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

```
vec1         <- c(3, 5, 8, 10, 9)
vec2         <- c(1, 10, 7, 3, 5)

diff <- vec1 - vec2
diff
```

```
[1]  2 -5  1  7  4
```
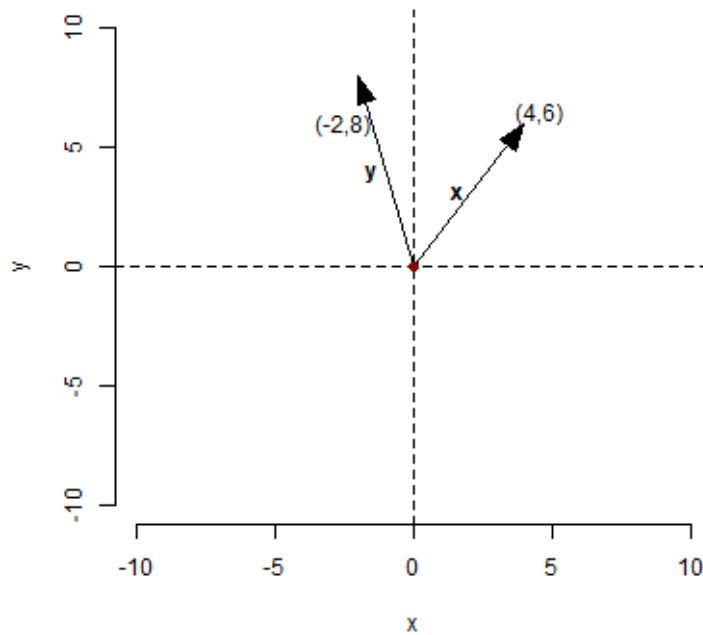
```
sqrt(diff%*%diff)
```

```
        [,1]
[1,] 9.746794
```

The angle between two vectors

Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in $\mathbb{R}^2$.

$$\mathbf{x} = \begin{bmatrix} 4 \\ 6 \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} -2 \\ 8 \end{bmatrix},$$

We can compute the angle between two vectors (θ) using the following formula.

$$\theta = cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|}\right)$$

```r
vec1          <- c(4,6)
vec2          <- c(-2,8)

vec1.norm <- sqrt(vec1%*%vec1)
vec2.norm <- sqrt(vec2%*%vec2)
dot.prod  <- vec1%*%vec2

# acos() is a function to compute the inverse of cosinus function
# returns radian

angle <- acos(dot.prod/(vec1.norm*vec2.norm))
angle
```

```
          [,1]
[1,] 0.8329813
```

```r
# Convert from radian to degree

angle*(180/pi)
```

```
          [,1]
[1,] 47.72631
```

Let's see if we can recover the angle between two orthogonal vectors (90 degree) using this formula.
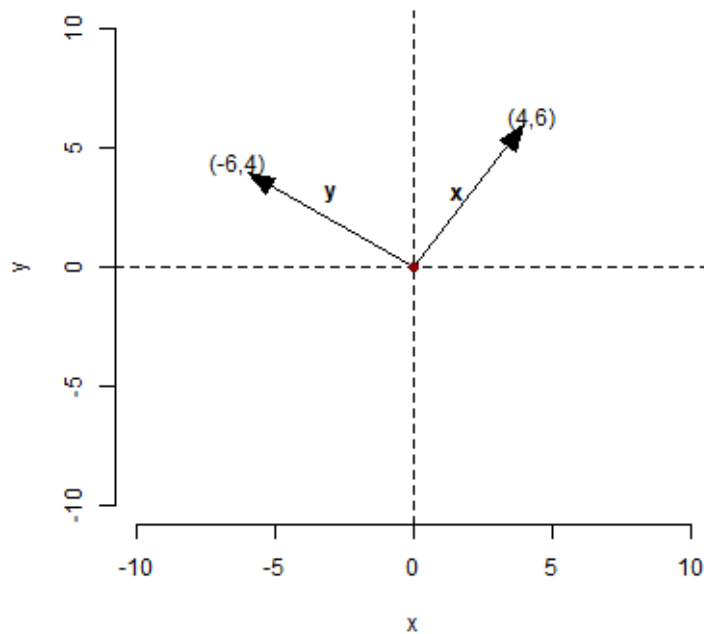

```
vec1          <- c(4,6)
vec2          <- c(-6,4)

vec1.norm <- sqrt(vec1%*%vec1)
vec2.norm <- sqrt(vec2%*%vec2)
dot.prod  <- vec1%*%vec2

# acos() is a function to compute the inverse of cosinus function
# returns radian

angle <- acos(dot.prod/(vec1.norm*vec2.norm))
angle
```

```
        [,1]
[1,] 1.570796
```

```
# Convert from radian to degree

angle*(180/pi)
```

```
      [,1]
[1,]    90
```

Note that this formula extends to any dimension although it is tough to visualize or conceptualize.

```
vec1          <- c(3, 5, 8, 10, 9)
vec2          <- c(1, 10, 7, 3, 5)

vec1.norm <- sqrt(vec1%*%vec1)
vec2.norm <- sqrt(vec2%*%vec2)
dot.prod  <- vec1%*%vec2

# acos() is a function to compute the inverse of cosinus function
# returns radian

angle <- acos(dot.prod/(vec1.norm*vec2.norm))
angle
```

```
          [,1]
[1,] 0.623063
```

```
# Convert from radian to degree

angle*(180/pi)
```

```
          [,1]
[1,] 35.69888
```

Unit vectors

A vector with a norm equal to 1 is called a unit vector. You can always find a unit vector in the direction of any non-zero vector if you divide the vector by its length. This is also called normalizing.

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

$$\|\hat{\mathbf{x}}\| = 1$$

```
vec1          <- c(3, 5, 8, 10, 9)

# Let's normalize this vector and find a unit vector in the same direction

l <- as.numeric(sqrt(vec1%*%vec1))
l
```

```
[1] 16.70329
```

```
unit <- vec1/l
unit
```

```
[1] 0.1796053 0.2993422 0.4789475 0.5986843 0.5388159
```

```
# Check if the length of the unit vector is indeed equal to 1

sqrt(unit%*%unit)
```

```
     [,1]
[1,]    1
```

Standard unit vectors

Standard unit vectors are specific unit vectors with length one and has 1 in the nth position and 0 elsewhere. For instance, in $\mathbb{R}^2$, the standard unit vectors are

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Similarly, in $\mathbb{R}^3$, the standard unit vectors can be written as

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

In its most general form, in $\mathbb{R}^n$, the standard unit vectors can be written as

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ . \\ . \\ . \\ 0 \end{bmatrix} ....., e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 1 \end{bmatrix}.$$

The shortest distance between a vector and an hyperplane

The shortest distance from a vector $\mathbf{u}$ to any point on the hyperplane

$$\mathbf{m} \cdot \mathbf{H} + b = 0$$

is equal to

$$\frac{|\mathbf{u} \cdot \mathbf{m} + b|}{\|\mathbf{m}\|}.$$

Let's try to understand this in a 2-dimensional space. In a two dimensional space, a hyperplane is a line and can be specified using a linear equation. For instance, suppose one wants to find the shortest distance between a point with the coordinates of (3,9) and a line with the equation of

$$3x + y + 4 = 0.$$
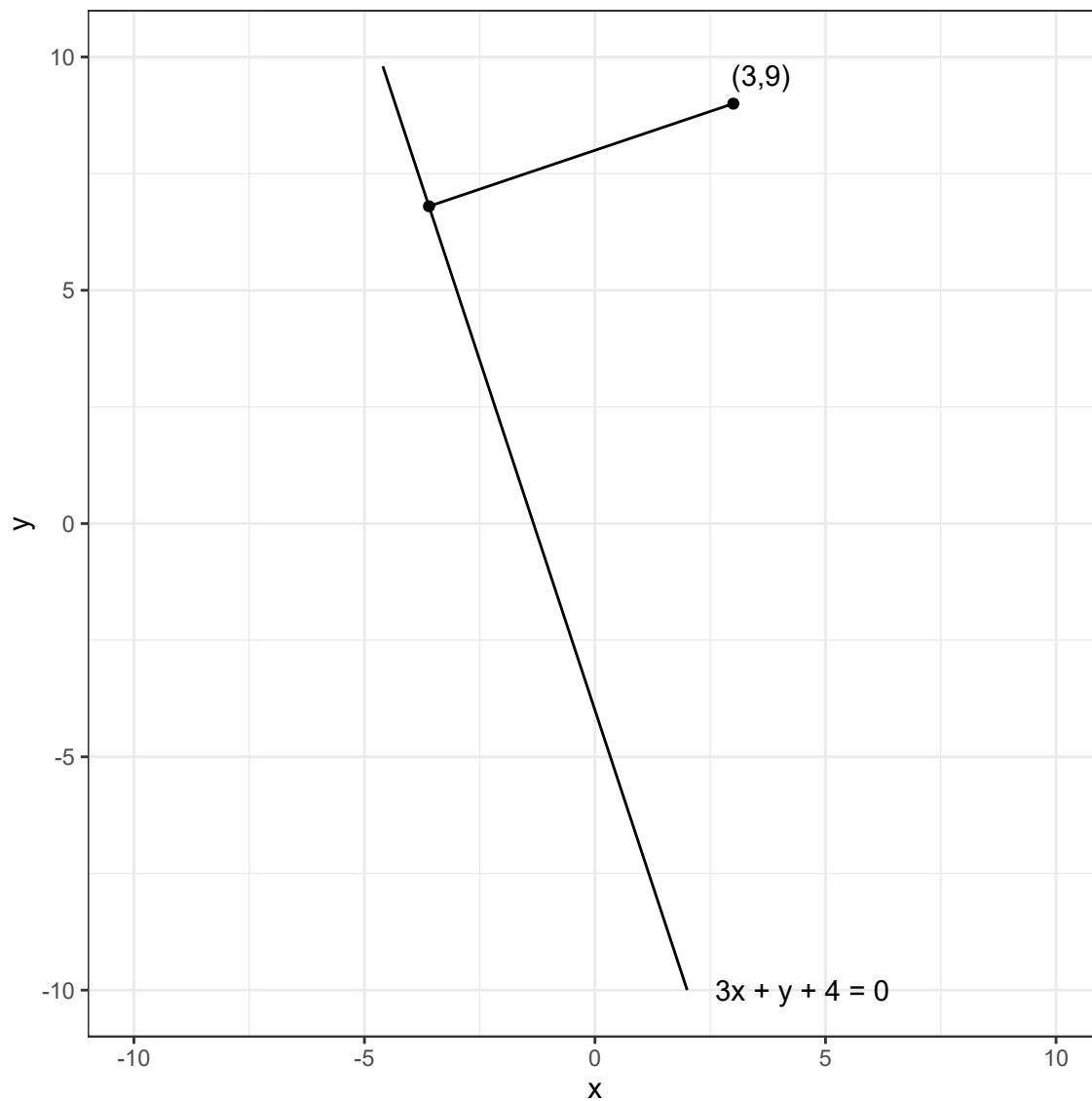
First, we write the line equation (hyperplane in 2D) in the form of

$$\mathbf{m} \cdot \mathbf{H} + b = 0$$

.

$$\begin{bmatrix} 3 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + 4 = 0$$

So, in this case, $\mathbf{m}$, b, and $\mathbf{u}$ can be specified as the following:

$$\mathbf{m} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$b = 4$$

$$\mathbf{u} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

The distance can be found using the formula above:

```r
m <- c(3, 1)
b <- 4
u <- c(3,9)

abs((u%*%m + b))/sqrt(m%*%m)
```

```
          [,1]
[1,]  6.957011
```

We can generalize this concept to a 3-dimensional space. suppose one wants to find the shortest distance between a point with the coordinates of (-30,15,25) and a hyperplane with the equation of

$$3x + y + 2z + 8 = 0.$$
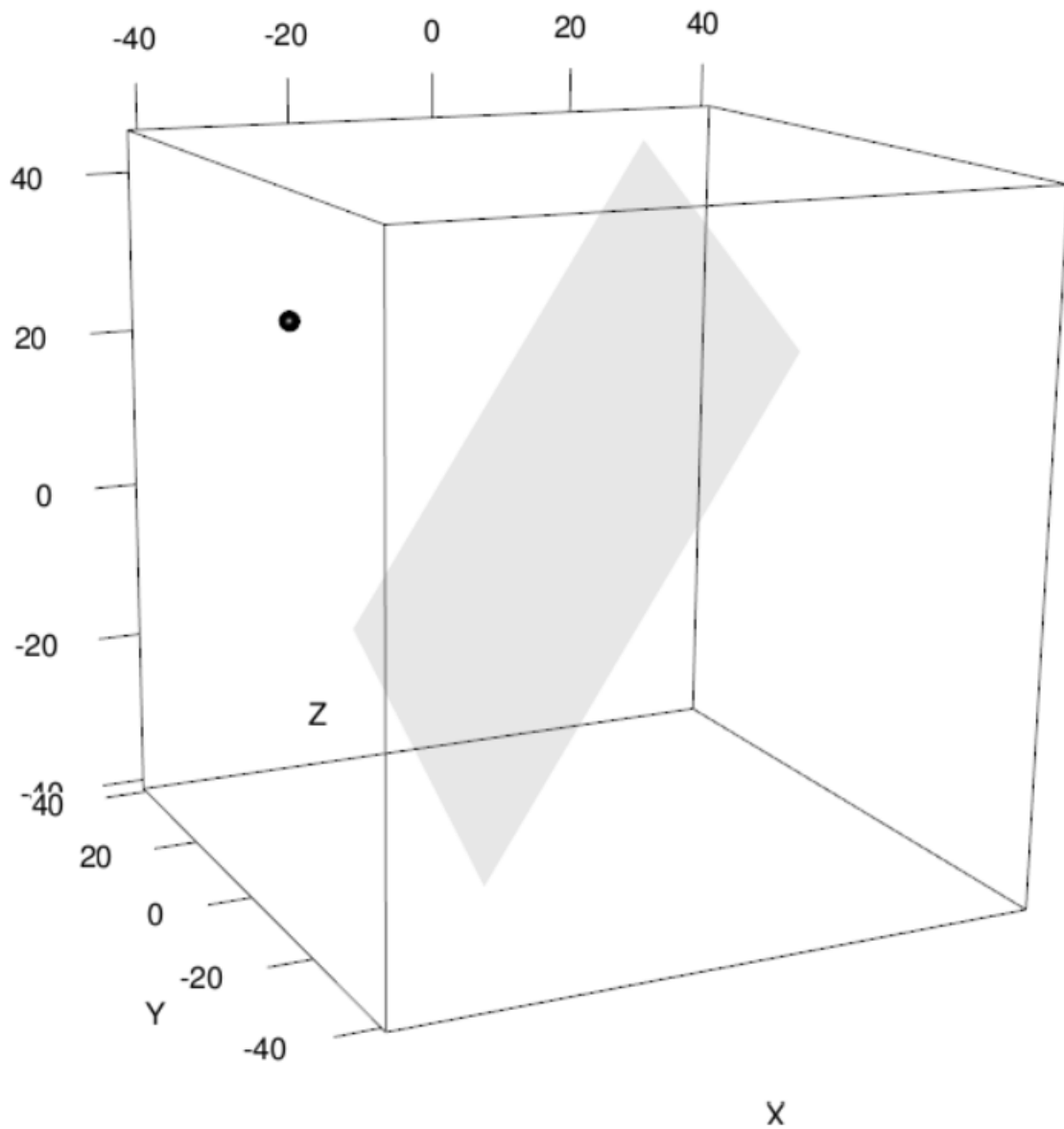
Let's write the hyperplane equation in the form of

$$\mathbf{m} \cdot \mathbf{H} + b = 0$$

.

$$\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + 8 = 0$$

Similarly, $\mathbf{m}$, b, and $\mathbf{u}$ can be specified as the following:

$$\mathbf{m} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

16

$$b = 8$$

$$\mathbf{u} = \begin{bmatrix} -30 \\ 15 \\ 25 \end{bmatrix}$$

```r
m <- c(3, 1, 2)
b <- 8
u <- c(-30, 15, 25)

abs((u%*%m + b))/sqrt(m%*%m)
```

```
          [,1]
[1,] 4.543441
```

This formula generalizes to larger dimensions.

Cauchy–Schwarz inequality

Let $\mathbf{x}$ and $\mathbf{y}$ be two non-zero vectors in $\mathbb{R}^n$. Then, the Cauchy–Schwarz inequality shows that the dot product is less than or equal to the product of norms of the vectors.

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$$

Minkowski inequality

Let $\mathbf{x}$ and $\mathbf{y}$ be two non-zero vectors in $\mathbb{R}^n$. Then, Minkowsky inequality shows that the norm of the addition of vectors is less than the addition of norms of each vector.

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

Linear Independence

The vectors $\mathbf{x_1}$, $\mathbf{x_2}$, ... , $\mathbf{x_k}$ in $\mathbb{R}^n$ are linearly independent when each vector is not a scalar multiples of each other. If a set of vectors are linearly independent, the only way the following holds is when all the scalars $(k_1, k_2, \ldots, k_n)$ are equal to zero.

$$k_1\mathbf{x_1} + k_2\mathbf{x_2} + ... + k_n\mathbf{x_n} = \mathbf{0}$$

For instance, consider the following three vectors.

$$\mathbf{x_1} = \begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}, \mathbf{x_2} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \mathbf{x_3} = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}.$$

These three vectors are NOT linearly independent because we can show that $\mathbf{x_3}$ can be obtained as a scalar multiples of $\mathbf{x_1}$ and $\mathbf{x_2}$.

$$\mathbf{x_3} = 2\mathbf{x_1} - 2\mathbf{x_2}$$

$$\begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix} = 2\begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

In another words, because the scalars that satisfy the following equation are not all equal to zero, $k_1 = 2$, $k_2 = -2$, and $k_3 = -1$.

$$k_1\mathbf{x_1} + k_2\mathbf{x_2} + k_3\mathbf{x_3} = \mathbf{0}.$$

$$2\begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Matrices

A matrix is a set of numbers arranged in rows and columns to form a rectangular array. A matrix has two dimensions given by number of rows and number of columns. For instance, the following is a 3 x 4 matrix.

$$\mathbf{A} = \begin{bmatrix} 3 & 6 & 8 & 1 \\ 5 & 9 & 3 & 0 \\ 8 & 4 & 2 & 1 \end{bmatrix}$$

Note that the columns of a matrix are called the column vectors and rows of a matrix are called row vectors. For instance, the matrix above has four column vectors,

$$\mathbf{A_{.1}} = \begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}, \mathbf{A_{.2}} = \begin{bmatrix} 6 \\ 9 \\ 4 \end{bmatrix}, \mathbf{A_{.3}} = \begin{bmatrix} 8 \\ 3 \\ 2 \end{bmatrix}, \mathbf{A_{.1}} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

and three row vectors,

$$\mathbf{A_{1.}} = \begin{bmatrix} 3 & 6 & 8 & 1 \end{bmatrix}, \mathbf{A_{2.}} = \begin{bmatrix} 5 & 9 & 3 & 0 \end{bmatrix}, \mathbf{A_{3.}} = \begin{bmatrix} 8 & 4 & 2 & 1 \end{bmatrix}.$$

Matrix Arithmetics

Zero matrix

A matrix with all elements equal to zero is called a zero matrix.

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & . & . & . & 0 \\ 0 & 0 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 0 \end{bmatrix}$$

Identity matrix

A matrix with diagonal elements equal to one and all other elements equal to zero is called an identity matrix.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 \end{bmatrix}$$

Scalar multiplication

When a matrix is multiplied by a scalar, the result is a matrix with the same size and each element of the matrix is multiplied with the same scalar.

$$\mathbf{A} = \begin{bmatrix} -1 & 5 & 4 \\ 3 & -2 & 0 \\ 9 & 8 & 12 \end{bmatrix}$$

$$3 \cdot \mathbf{A} = \begin{bmatrix} -3 & 15 & 12 \\ 9 & -6 & 0 \\ 27 & 24 & 36 \end{bmatrix}$$

```r
A <- matrix(data = c(-1,5,4,3,-2,0,9,8,12),nrow=3,ncol=3,byrow=TRUE)
A
```

```
     [,1] [,2] [,3]
[1,]   -1    5    4
[2,]    3   -2    0
[3,]    9    8   12
```

```r
3*A
```

```
     [,1] [,2] [,3]
[1,]   -3   15   12
[2,]    9   -6    0
[3,]   27   24   36
```

Below are some properties of scalar multiplication, and I will leave it to you if you want to verify them. Suppose $\mathbf{A}$ and $\mathbf{B}$ are both matrices with the same dimensions and c and k are two scalars.

- $(ck)\mathbf{A} = c(k\mathbf{A}) = k(c\mathbf{A})$

- $k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$

- $(c + k)\mathbf{A} = c\mathbf{A} + k\mathbf{A}$

Matrix addition

Two matrices with the same size can be added together, and the result is a matrix with the same size and each corresponding element of two matrices are added.

$$\mathbf{A} = \begin{bmatrix} -1 & 5 & 4 \\ 3 & -2 & 0 \\ 9 & 8 & 12 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 4 & 0 & 1 \\ -5 & 2 & -1 \\ 15 & 2 & 4 \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} (-1)+4 & 5+0 & 4+1 \\ 3+(-5) & (-2)+2 & 0+(-1) \\ 9+15 & 8+2 & 12+4 \end{bmatrix} = \begin{bmatrix} 3 & 5 & 5 \\ -2 & 0 & -1 \\ 24 & 10 & 16 \end{bmatrix}$$

```r
A <- matrix(data = c(-1,5,4,3,-2,0,9,8,12),nrow=3,ncol=3,byrow=TRUE)
A
```

```
     [,1] [,2] [,3]
[1,]   -1    5    4
[2,]    3   -2    0
[3,]    9    8   12
```

```r
B <- matrix(data = c(4,0,1,-5,2,-1,15,2,4),nrow=3,ncol=3,byrow=TRUE)
B
```

```
     [,1] [,2] [,3]
[1,]    4    0    1
[2,]   -5    2   -1
[3,]   15    2    4
```

```r
A+B
```

```
     [,1] [,2] [,3]
[1,]    3    5    5
[2,]   -2    0   -1
[3,]   24   10   16
```

Below are some properties of matrix addition, and I will leave it to you if you want to verify them. Suppose $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are matrices with the same dimensions.

- $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

- $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$

- There is a zero matrix with the same dimensions that always satisfies the following $\mathbf{A} + \mathbf{0} = \mathbf{A}$

- There is always a matrix called additive inverse such that $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$

Transpose of a matrix

The transpose of a matrix is when you rotate the rows and columns of a matrix so that the old rows become the new columns and old columns become the new rows. The transpose of a matrix $\mathbf{A}$ is denoted as $\mathbf{A}^T$. Below is an example of a transpose of a matrix. If a matrix has dimensions of r rows and c columns, the transpose of the matrix will have dimensions of c rows and r columns.

$$\mathbf{A} = \begin{bmatrix} -1 & 5 & 4 & 0 \\ 3 & -2 & 0 & 1 \end{bmatrix}$$

$$\mathbf{A}^T = \begin{bmatrix} -1 & 3 \\ 5 & -2 \\ 4 & 0 \\ 0 & 1 \end{bmatrix}$$

In R, you can use the `t()` function to transpose a matrix.

```
A <- matrix(data = c(-1,5,4,0,3,-2,0,1),nrow=2,ncol=4,byrow=TRUE)
A
```

```
     [,1] [,2] [,3] [,4]
[1,]   -1    5    4    0
[2,]    3   -2    0    1
```

```
dim(A)
```

```
[1] 2 4
```

```
t(A)
```

```
     [,1] [,2]
[1,]   -1    3
[2,]    5   -2
[3,]    4    0
[4,]    0    1
```

```
dim(t(A))
```

```
[1] 4 2
```

Sometimes, you will see that a column vector is written as transpose of a row vector to save space.

$$\mathbf{B} = \begin{bmatrix} -1 \\ 5 \\ 4 \\ 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} -1 & 5 & 4 & 0 \end{bmatrix}^T$$

Below are some properties of matrix matrix transpose. Suppose $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{C}$ are matrices with the appropriate sizes.

- $(\mathbf{A}^T)^T = \mathbf{A}$

- $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

- $(\mathbf{A} \times \mathbf{B})^T = \mathbf{B}^T \times \mathbf{A}^T$

- $(k\mathbf{A})^T = k\mathbf{A}^T$

Matrix multiplication

In order to be able to multiply two matrices, their dimensions should match. The number of columns of the left-hand matrix must be equal to the number of rows of the right-hand matrix. If the left-hand matrix has dimensions of $r \times m$ and the right-hand matrix has dimensions of $m \times n$, then the result is always a matrix with dimensions of $r \times n$.

Matrix multiplication doesn't happen how it sounds. We don't just multiply the corresponding elements (as it is done for addition). Instead, each row vector of the left-hand matrix is entered to a dot product with each column vector of the right-hand matrix.

Consider the following two matrices $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{A} = \begin{bmatrix} -1 & 5 & 4 & 0 \\ 3 & -2 & 0 & 1 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 4 & 0 \\ -5 & 2 \\ 15 & 2 \\ 0 & -3 \end{bmatrix}$$

$\mathbf{A}$ is a 2 x 4 matrix and $\mathbf{B}$ is a 4 x 2 matrix. When we multiply these two matrices, the result will be a 2 x 2 matrix. In other words, $\mathbf{A}$ has two row vectors and $\mathbf{B}$ has two column vectors, so we will do 4 different dot products in total.

1. The first row of $\mathbf{A}$ and the first column of $\mathbf{B}$

$$\begin{bmatrix} -1 & 5 & 4 & 0 \end{bmatrix}^T \cdot \begin{bmatrix} 4 \\ -5 \\ 15 \\ 0 \end{bmatrix} = (-1 \times 4) + (5 \times -5) + (4 \times 15) + (0 \times 0) = 31$$

2. The second row of $\mathbf{A}$ and the first column of $\mathbf{B}$

$$\begin{bmatrix} 3 & -2 & 0 & 1 \end{bmatrix}^T \cdot \begin{bmatrix} 4 \\ -5 \\ 15 \\ 0 \end{bmatrix} = (3 \times 4) + (-2 \times -5) + (0 \times 15) + (1 \times 0) = 22$$

3. The first row of $\mathbf{A}$ and the second column of $\mathbf{B}$

$$\begin{bmatrix} -1 & 5 & 4 & 0 \end{bmatrix}^T \cdot \begin{bmatrix} 0 \\ 2 \\ 2 \\ -3 \end{bmatrix} = (-1 \times 0) + (5 \times 2) + (4 \times 2) + (0 \times -3) = 18$$

2. The second row of **A** and the second column of **B**

$$\begin{bmatrix} 3 & -2 & 0 & 1 \end{bmatrix}^T \cdot \begin{bmatrix} 0 \\ 2 \\ 2 \\ -3 \end{bmatrix} = (3 \times 0) + (-2 \times 2) + (0 \times 2) + (1 \times -3) = -7$$

So, the final result would be:

$$\mathbf{A} \times \mathbf{B} = \begin{bmatrix} 31 & 18 \\ 22 & -7 \end{bmatrix}$$

```r
A <- matrix(data = c(-1,5,4,0,3,-2,0,1),nrow=2,ncol=4,byrow=TRUE)
A
```

```
     [,1] [,2] [,3] [,4]
[1,]   -1    5    4    0
[2,]    3   -2    0    1
```

```r
B <- matrix(data = c(4,0,-5,2,15,2,0,-3),nrow=4,ncol=2,byrow=TRUE)
B
```

```
     [,1] [,2]
[1,]    4    0
[2,]   -5    2
[3,]   15    2
[4,]    0   -3
```

```r
A%*%B
```

```
     [,1] [,2]
[1,]   31   18
[2,]   22   -7
```

Below are some important properties of matrix multiplication. Suppose **A**, **B**, **C** are matrices with appropriate dimensions:

- Matrix multiplication is NOT commutative, - $\mathbf{A} \times \mathbf{B} \neq \mathbf{B} \times \mathbf{A}$

- $\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) + (\mathbf{A} \times \mathbf{C})$

- $(\mathbf{A} + \mathbf{B}) \times \mathbf{C} = (\mathbf{A} \times \mathbf{C}) + (\mathbf{B} \times \mathbf{C})$

- $(\mathbf{A} \times \mathbf{0}) = \mathbf{0}$

- $\mathbf{A} \times \mathbf{A} = \mathbf{A}^2$

- $\mathbf{A}^0 = \mathbf{I}$

- $\mathbf{A} \times \mathbf{I} = \mathbf{A}$

Inverse of a matrix

There is no matrix division. The closes concept of division in matrix algebra is an inverse of a matrix. The inverse of a matrix $\mathbf{A}$ is denoted as $\mathbf{A}^{-1}$. Suppose $\mathbf{A}$ is a square matrix, the inverse of $\mathbf{A}$ satisfies the following equation:

$$\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{I}$$

In R, you can use the `solve()` function to find the inverse of a matrix.

```
A <- matrix(data = c(1,2,3,4,0,5,7,9,-2,4,-8,5,10,6,6,7),nrow=4,ncol=4,byrow=TRUE)
A
```

```
     [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    0    5    7    9
[3,]   -2    4   -8    5
[4,]   10    6    6    7
```

```
# Find the inverse of A

B <- solve(A)

B
```

```
           [,1]       [,2]         [,3]        [,4]
[1,]  0.4793388 -0.2520661 -0.002066116  0.05165289
[2,] -2.1900826  0.7809917  0.030991736  0.22520661
[3,] -0.3057851  0.1694215 -0.080578512  0.01446281
[4,]  1.4545455 -0.4545455  0.045454545 -0.13636364
```

```
# Show that the A times inverse of A is an identity matrix

round(A%*%B,3)
```

```
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
```

Determinant of a matrix

Every square matrix has a unique value associated with it. This value can be used to examine whether or not a matrix has an inverse. If the determinant of a matrix is equal or close to zero, there is no inverse. In R, you can use the `det()` function to compute the determinant of a matrix. Let's check the two examples from the previous section.

```
A <- matrix(data = c(1,1,2,-1,3,-5,2,-2,7),nrow=3,ncol=3,byrow=TRUE)
A
     [,1] [,2] [,3]
```

```
[1,]    1    1    2
[2,]   -1    3   -5
[3,]    2   -2    7

det(A)
[1] 0

  # equals to zero so there is no inverse of A
```

```
A <- matrix(data = c(1,2,3,4,0,5,7,9,-2,4,-8,5,10,6,6,7),nrow=4,ncol=4,byrow=TRUE)
A
     [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    0    5    7    9
[3,]   -2    4   -8    5
[4,]   10    6    6    7

det(A)
[1] 484

  # different than zero, so there is inverse of A
```

If a matrix has no inverse, it is called a non-ivertible or singular matrix.

Rank of a matrix

The rank of a matrix is the maximal number of linearly independent columns in the matrix. You can use the `rankMatrix()` function from the `Matrix` package to find the rank of a matrix. Consider the following matrix with three columns.

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 4 \\ 5 & 2 & 6 \\ 8 & 4 & 8 \end{bmatrix}.$$

```
require(Matrix)

A <- matrix(data = c(3,1,4,5,2,6,8,4,8),nrow=3,ncol=3,byrow=TRUE)
A
     [,1] [,2] [,3]
[1,]    3    1    4
[2,]    5    2    6
[3,]    8    4    8

rankMatrix(A)
[1] 2
attr(,"method")
[1] "tolNorm2"
attr(,"useGrad")
[1] FALSE
attr(,"tol")
[1] 0.000000000000000006661338
```

In this case, this returns 2 as the rank of this matrix. The reason is that the third column of this matrix can indeed be obtained from the first two columns, see the section on Linear Independence.

Trace of a matrix

The trace of a matrix $\mathbf{A}$ is the sum of its diagonal elements and is denoted as $tr(\mathbf{A})$.

```
require(Matrix)

A <- matrix(data = c(3,1,4,5,2,6,8,4,8),nrow=3,ncol=3,byrow=TRUE)
A
     [,1] [,2] [,3]
[1,]    3    1    4
[2,]    5    2    6
[3,]    8    4    8

sum(diag(A))
[1] 13
```

Norm of a matrix

The norm of a matrix $\mathbf{A}$ can be calculated using

$$||\mathbf{A}|| = tr(\mathbf{A}^T \times \mathbf{A})$$

```
require(Matrix)

A <- matrix(data = c(3,1,4,5,2,6,8,4,8),nrow=3,ncol=3,byrow=TRUE)
A
     [,1] [,2] [,3]
[1,]    3    1    4
[2,]    5    2    6
[3,]    8    4    8

t(A)%*%A
     [,1] [,2] [,3]
[1,]   98   45  106
[2,]   45   21   48
[3,]  106   48  116

sum(diag(t(A)%*%A))
[1] 235
```

Solving Sysyem of Linear Equations

A finite number of linear equations with a finite number of unknowns is known as a system of linear equations. You can easily solve these systems based on what you learned up to this point. Suppose, you have the following system of linear equations with 4 unknowns.

$$x + 3y + z - 4k = 7$$

$$3x - y + 2z + k = 12$$

$$2x - 4y + 5z - 10k = 1$$

$$x + y + z + k = 5$$

What values of x, y, z, and k satisfy these equations?

When we try to solve a system of linear equations, we try to write them in the following format:

$$\mathbf{A} \times \mathbf{x} = \mathbf{b},$$

where $\mathbf{A}$ represents a matrix of the coefficients related to the unknowns, $\mathbf{x}$ represents a column vector of unknowns, and $\mathbf{b}$ represents the outcome of these equations. When we multiple each side of this equation with the inverse of $\mathbf{A}$, we can solve these equations for $\mathbf{x}$.

$$\mathbf{A}^{-1} \times \mathbf{A} \times \mathbf{x} = \mathbf{A}^{-1} \times \mathbf{b},$$

$$(\mathbf{A}^{-1} \times \mathbf{A}) \times \mathbf{x} = \mathbf{A}^{-1} \times \mathbf{b},$$

$$\mathbf{I} \times \mathbf{x} = \mathbf{A}^{-1} \times \mathbf{b},$$

$$\mathbf{x} = \mathbf{A}^{-1} \times \mathbf{b},$$

Let's do this for the example above.

$$\begin{bmatrix} 1 & 3 & 1 & -4 \\ 3 & -1 & 2 & 1 \\ 2 & -4 & 5 & -10 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ z \\ k \end{bmatrix} = \begin{bmatrix} 7 \\ 12 \\ 1 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \\ k \end{bmatrix} = \begin{bmatrix} 1 & 3 & 1 & -4 \\ 3 & -1 & 2 & 1 \\ 2 & -4 & 5 & -10 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \times \begin{bmatrix} 7 \\ 12 \\ 1 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \\ k \end{bmatrix} = \begin{bmatrix} 0.29 & 0.60 & -0.13 & -0.80 \\ 0.14 & -.12 & -0.04 & 0.29 \\ -0.29 & -0.43 & 0.19 & 1.19 \\ -0.14 & -0.05 & -0.02 & 0.32 \end{bmatrix} \times \begin{bmatrix} 7 \\ 12 \\ 1 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \\ k \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

```r
A <- matrix(data = c(1,3,1,-4,3,-1,2,1,2,-4,5,-10,1,1,1,1),nrow=4,ncol=4,byrow=TRUE)
A
```

```
      [,1] [,2] [,3] [,4]
[1,]     1    3    1   -4
[2,]     3   -1    2    1
[3,]     2   -4    5  -10
[4,]     1    1    1    1
```

```r
b <- c(7,12,1,5)
```

```r
# Find the solution
```

```r
x <- solve(A) %*% b
```

```r
x
```

```
      [,1]
[1,]     5
[2,]     1
[3,]    -1
[4,]     0
```

With some simple matrix arithmetic, we can easily solve a system of linear equations regardless of its size as long as you have enough data.

Note that there is not always a solution for a system of linear equations. Sometimes, a solution to one of the equations is inconsistent with a solution of another equation in the system. In such situations, there is no solution to the system and these systems are called inconsistent. See the following example. There is no solution because there is no inverse of **A**. R will give you a warning message indicating that the system is singular.

$$x + y + 2z = 3$$
$$-x + 3y - 5z = 7$$
$$2x - 2y + 7z = 1$$

```r
A <- matrix(data = c(1,1,2,-1,3,-5,2,-2,7),nrow=3,ncol=3,byrow=TRUE)
A
```

```
      [,1] [,2] [,3]
[1,]     1    1    2
[2,]    -1    3   -5
[3,]     2   -2    7
```

```r
b <- c(3,7,1)
```

```r
solve(A)
```

```
Error in solve.default(A): Lapack routine dgesv: system is exactly singular: U[3,3] = 0
```

Also, if there are fewer equations than the number unknowns, there is an infinite number of solution to a system of linear equation. See the following example. There is no solution because there is no inverse of **A**. R will give you a warning message indicating that the system is singular.

On the other hand, this system has infinite number of solutions because there are only two equations and three unknowns.

$$x + y + 2z = 3$$

$$-x + 3y - 5z = 7$$