matrixNotation

September 4, 2019

1 matrix notation for statistics

Matrix and vector notation comes from the branch of mathematics called Linear Algebra. Matrices and vectors are used in essentially all statistical disciplines, and are a key to understanding theoretical and empirical properties of our statistical models. We'll define a vector and matrix, and how they interact in the context of statistics.

Our end goal will be to understand how linear regression can be represented using vectors and matrices, understanding this equation

$$Y \sim \mathcal{N}(X\beta, \sigma^2)$$

First, we'll define vectors and understand how they can represent data and points in n-dimensional space. Next we'll define matrices and different ways of conceptualizing them. After defining vectors and matrices separately, we'll learn about how they interact with one another: vector times vector, vector times matrix, and matrix times matrix. Finally, we'll frame linear regression in terms of matrices and vectors.

1.1 vector

A vector is an n-dimensional collection of real numbers. Vectors are represented as a stack of these n numbers, surrounded by brackets. For example,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

is a vector with n elements.

A specific example of a vector with values is

$$h = \begin{bmatrix} 34 \\ 22.2 \\ 83/2 \end{bmatrix}$$

h is a vector of length 3. The first element is 34, second 22.2, and third 83/2.

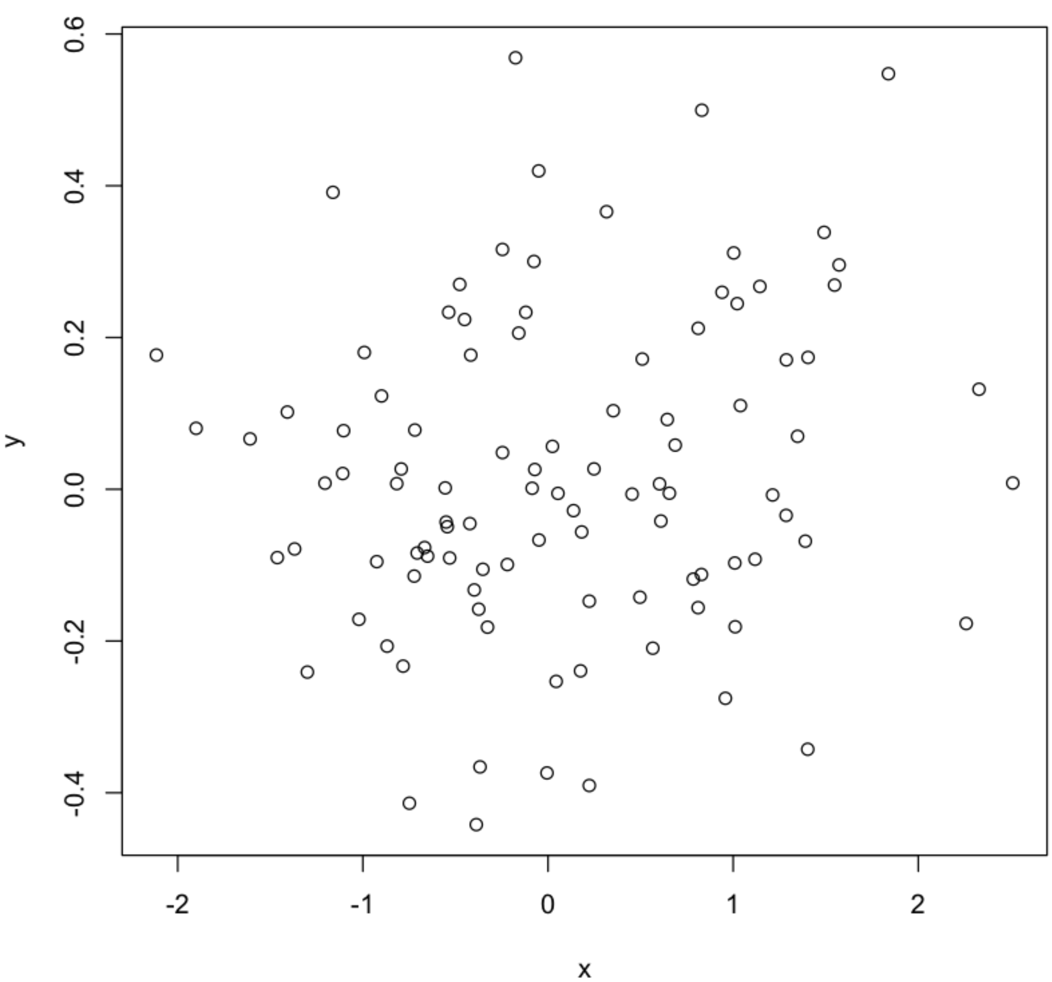
Vectors generalize the number line. A n-dimensional vector represents a single point in a n-dimensional space of real numbers. Real numbers can be thought of as 1-dimensional vectors. For example the expression $r = 45.456$ could be represented as

$$r = \begin{bmatrix} 45.456 \end{bmatrix},$$

a 1-dimensional vector with a single value 45.456.

Vectors can help us gain intuition about how statistical models are fit. For example, scatterplots can be thought of as a collection of 2-dimensional vectors

[1]:
```
x <- rnorm(100,0,1)
y <- rnorm(100,0,0.2)
plot(x,y)
```



Each point on the scatter plot above can be represented as a "tuple" (x, y) or viewed as a single 2-dimensional vector.

1.2 Matrix

A matrix is a collection of vectors and has two dimensions written with square brackets, for example

$$M = \begin{bmatrix} 34 & 55 \\ 22.2 & 0.9 \\ 83/2 & 76.3 \end{bmatrix}$$

We'll use capital letters to denote matrices, like this

$$X = \begin{bmatrix} 4 & 55 \\ 1/3 & 1.9 \\ 0.88 & 99.3 \end{bmatrix}$$

Matrices can represent different concepts in statistics. They can represent P variables collected from a set of N observations. For example, height, weight, BMI, and CKMB enzyme levels for patients at risk for a myocardial infarction (heart attack).

In this example, each row would represent a patient, and the columns would represent height (inches), weight (kg), and CKMB levels (IU/L). Our matrix can be constructed like

$$D = \begin{bmatrix} 60 & 77 & 8 \\ 54 & 50 & 18 \\ 63 & 90 & 4 \end{bmatrix}$$

1.3 vector times vector (the inner product)

Consider two vectors a and b. Each is length 4, representing a single point in a 4-dimensional Real space.

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}; b = \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$$

Then we define the **inner product** of these two vectors as

$$
\begin{aligned}
a'b &= \sum_{i=1}^{4} a_i \times b_i \\
&= 1 \times 5 + 2 \times 6 + 3 \times 7 + 4 \times 8 \\
&= 70
\end{aligned}
$$

where the $'$ symbol stands for transpose.

The transpose "flips" the rows and columns of a vector or matrix. If

$$
a = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}
$$

a vector with 4 rows and 1 column then a' is a vector with 1 row and 4 columns

$$
a = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}
$$

1.4 matrix times matrix

Consider two matrices A and B. A has 2 rows and 3 columns. B has 3 rows and 5 columns.

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

$$
B = \begin{bmatrix} 7 & 8 & 9 & 10 & 11 \\ 12 & 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 & 21 \end{bmatrix}
$$

Then we define the (i, j) element of the product of matrix A and B as

$$
AB(i, j) = \sum_{r=1}^{3} a(i, r) b(r, j)
$$

Another way to look at this matrix multiplication product is by using an inner product. Redefine the matrix A in terms of two "row" vectors and B in terms of 5 "column" vectors.

$$
\begin{aligned}
A &= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \\
&= \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
a_1 &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\
a_2 &= \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}
\end{aligned}
$$

and

$$
B = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 \end{bmatrix}
$$

where

$$
b_1 = \begin{bmatrix} 7 \\ 12 \\ 17 \end{bmatrix}
$$

and so on. Then we can redefine the (i, j) element of A times B as an inner product

$$
\begin{aligned}
AB(i, j) &= \sum_{r=1}^{3} a(i, r) b(r, j) \\
&= a_i' b_j
\end{aligned}
$$

1.5 matrix times vector

Consider a vector β and matrix X. The product is a vector, where the (i^{th}) element equals

$$
X\beta(i) = \sum_j X(i, j)\beta_j
$$

or the inner product

$$
X\beta(i) = x_i'\beta
$$

where x_i is the i^{th} row vector of the matrix X

1.6 matrix notation for simple linear regression

The above terminology and methods can allow us to express linear regression as an operation on matrices and vectors.

1.6.1 model form (system of equations)

Simple linear regression relates a variable x to a variable y by the formula

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x + \epsilon \\
\epsilon &\sim N(0, \sigma^2)
\end{aligned}
$$

A simple linear regression equation states any value y can be probabilistically predicted or explained by knowing the corresponding value x, β_0, β_1, and the variance σ^2 . Given a list of n values for x, a value for β_0 and a value for β_1, we can rewrite the above equation in matrix notation.

$$y = X\beta + \epsilon$$

where y is a vector of length n,

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

β is a vector containing β_0 and β_1,

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix},$$

and

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

is a matrix; the first column all ones and the second column the values of x. Epsilon is a vector of length n

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

every value following a Normal distribution with variance σ^2.

Written this way, we can define a structure or functional form for all values of y and x at once. The functional form does not look that different from our original $y = \beta_0 + \beta_1 x + \epsilon$, and is even more compact.

1.6.2 probability form

The equation

$$
\begin{aligned}
y &= X\beta + \epsilon \\
\epsilon &\sim N(0, \sigma^2)
\end{aligned}
$$

explains the probabilistic, and linear relationship, between x and y. But I believe we can do better. The above equation can confuse the non-statistically inclined. Often, they may focus on the linear relationship, forgetting the probabilistic component because of how it is written. The ϵ appears like an after thought.

To more clearly communicate the fact that the relationship between y and x is probabilistic, we can write

$$y \sim N(X\beta, \sigma^2),$$

read y is Normally distributed with mean $X\beta$ and variance σ^2. This expression, called the probabilistic form, more clearly communicates that the relationship between y and x is not deterministic, and allowed to fluctuate around a mean that depends on x and on β.

```
[ ]:
```