Linear algebra explained in four pages

Excerpt from the NO BULLSHIT GUIDE TO LINEAR ALGEBRA by Ivan Savov

Abstract—**This document will review the fundamental ideas of linear algebra. We will learn about matrices, matrix operations, linear transformations and discuss both the theoretical and computational aspects of linear algebra. The tools of linear algebra open the gateway to the study of more advanced mathematics. A lot of *knowledge buzz* awaits you if you choose to follow the path of *understanding*, instead of trying to memorize a bunch of formulas.**

I. INTRODUCTION

Linear algebra is the math of vectors and matrices. Let n be a positive integer and let $\mathbb{R}$ denote the set of real numbers, then $\mathbb{R}^n$ is the set of all n-tuples of real numbers. A vector $\vec{v} \in \mathbb{R}^n$ is an n-tuple of real numbers. The notation "$\in S$" is read "element of S." For example, consider a vector that has three components:

$$\vec{v} = (v_1, v_2, v_3) \ \in \ (\mathbb{R}, \mathbb{R}, \mathbb{R}) \equiv \mathbb{R}^3.$$

A matrix $A \in \mathbb{R}^{m\times n}$ is a rectangular array of real numbers with m rows and n columns. For example, a 3×2 matrix looks like this:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \in \begin{bmatrix} \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} \end{bmatrix} \equiv \mathbb{R}^{3\times 2}.$$

The purpose of this document is to introduce you to the mathematical operations that we can perform on vectors and matrices and to give you a feel of the power of linear algebra. Many problems in science, business, and technology can be described in terms of vectors and matrices so it is important that you understand how to work with these.

Prerequisites

The only prerequisite for this tutorial is a basic understanding of high school math concepts[1] like numbers, variables, equations, and the fundamental arithmetic operations on real numbers: addition (denoted $+$), subtraction (denoted $-$), multiplication (denoted implicitly), and division (fractions).

You should also be familiar with *functions* that take real numbers as inputs and give real numbers as outputs, $f : \mathbb{R} \to \mathbb{R}$. Recall that, by definition, the *inverse function* f^{-1} *undoes* the effect of f. If you are given $f(x)$ and you want to find x, you can use the inverse function as follows: $f^{-1}(f(x)) = x$. For example, the function $f(x) = \ln(x)$ has the inverse $f^{-1}(x) = e^x$, and the inverse of $g(x) = \sqrt{x}$ is $g^{-1}(x) = x^2$.

II. DEFINITIONS

A. Vector operations

We now define the math operations for vectors. The operations we can perform on vectors $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ are: addition, subtraction, scaling, norm (length), dot product, and cross product:

$$\begin{aligned}
\vec{u} + \vec{v} &= (u_1 + v_1, u_2 + v_2, u_3 + v_3) \\
\vec{u} - \vec{v} &= (u_1 - v_1, u_2 - v_2, u_3 - v_3) \\
\alpha\vec{u} &= (\alpha u_1, \alpha u_2, \alpha u_3) \\
\|\vec{u}\| &= \sqrt{u_1^2 + u_2^2 + u_3^2} \\
\vec{u}\cdot\vec{v} &= u_1v_1 + u_2v_2 + u_3v_3 \\
\vec{u}\times\vec{v} &= (u_2v_3 - u_3v_2,\ u_3v_1 - u_1v_3,\ u_1v_2 - u_2v_1)
\end{aligned}$$

The dot product and the cross product of two vectors can also be described in terms of the angle θ between the two vectors. The formula for the dot product of the vectors is $\vec{u}\cdot\vec{v} = \|\vec{u}\|\|\vec{v}\|\cos\theta$. We say two vectors $\vec{u}$ and $\vec{v}$ are *orthogonal* if the angle between them is $90°$. The dot product of orthogonal vectors is zero: $\vec{u}\cdot\vec{v} = \|\vec{u}\|\|\vec{v}\|\cos(90°) = 0$.

The *norm* of the cross product is given by $\|\vec{u}\times\vec{v}\| = \|\vec{u}\|\|\vec{v}\|\sin\theta$. The cross product is not commutative: $\vec{u}\times\vec{v} \neq \vec{v}\times\vec{u}$, in fact $\vec{u}\times\vec{v} = -\vec{v}\times\vec{u}$.

B. Matrix operations

We denote by A the matrix as a whole and refer to its entries as a_{ij}. The mathematical operations defined for matrices are the following:

- addition (denoted $+$)
$$C = A + B \qquad \Leftrightarrow \qquad c_{ij} = a_{ij} + b_{ij}.$$
- subtraction (the inverse of addition)
- matrix product. The product of matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times\ell}$ is another matrix $C \in \mathbb{R}^{m\times\ell}$ given by the formula
$$C = AB \qquad \Leftrightarrow \qquad c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj},$$
$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix}$$
- matrix inverse (denoted A^{-1})
- matrix transpose (denoted $^{\mathsf{T}}$):
$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{bmatrix}.$$
- matrix trace: $\mathrm{Tr}[A] \equiv \sum_{i=1}^{n} a_{ii}$
- determinant (denoted $\det(A)$ or $|A|$)

Note that the matrix product is not a commutative operation: $AB \neq BA$.

C. Matrix-vector product

The matrix-vector product is an important special case of the matrix-matrix product. The product of a 3×2 matrix A and the 2×1 column vector $\vec{x}$ results in a 3×1 vector $\vec{y}$ given by:

$$\begin{aligned}
\vec{y} = A\vec{x} \qquad \Leftrightarrow \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \end{bmatrix} \\
&= x_1\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + x_2\begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \qquad \text{(C)} \\
&= \begin{bmatrix} (a_{11}, a_{12})\cdot\vec{x} \\ (a_{21}, a_{22})\cdot\vec{x} \\ (a_{31}, a_{32})\cdot\vec{x} \end{bmatrix}. \qquad \text{(R)}
\end{aligned}$$

There are two[2] fundamentally different yet equivalent ways to interpret the matrix-vector product. In the column picture, **(C)**, the multiplication of the matrix A by the vector $\vec{x}$ produces a **linear combination of the columns of the matrix**: $\vec{y} = A\vec{x} = x_1A_{[:,1]} + x_2A_{[:,2]}$, where $A_{[:,1]}$ and $A_{[:,2]}$ are the first and second columns of the matrix A.

In the row picture, **(R)**, multiplication of the matrix A by the vector $\vec{x}$ produces a column vector with coefficients equal to the **dot products of rows of the matrix** with the vector $\vec{x}$.

D. Linear transformations

The matrix-vector product is used to define the notion of a *linear transformation*, which is one of the key notions in the study of linear algebra. Multiplication by a matrix $A \in \mathbb{R}^{m\times n}$ can be thought of as computing a *linear transformation* T_A that takes n-vectors as inputs and produces m-vectors as outputs:

$$T_A : \mathbb{R}^n \to \mathbb{R}^m.$$

[1] A good textbook to (re)learn high school math is `minireference.com`

[2] For more info see the video of Prof. Strang's MIT lecture: `bit.ly/1OvmKcL`

Limits

Definitions

Precise Definition : We say $\lim_{x\to a} f(x) = L$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that whenever $0 < |x-a| < \delta$ then $|f(x) - L| < \varepsilon$.

"Working" Definition : We say $\lim_{x\to a} f(x) = L$ if we can make $f(x)$ as close to L as we want by taking x sufficiently close to a (on either side of a) without letting $x = a$.

Right hand limit : $\lim_{x\to a^+} f(x) = L$. This has the same definition as the limit except it requires $x > a$.

Left hand limit : $\lim_{x\to a^-} f(x) = L$. This has the same definition as the limit except it requires $x < a$.

Limit at Infinity : We say $\lim_{x\to\infty} f(x) = L$ if we can make $f(x)$ as close to L as we want by taking x large enough and positive.

There is a similar definition for $\lim_{x\to-\infty} f(x) = L$ except we require x large and negative.

Infinite Limit : We say $\lim_{x\to a} f(x) = \infty$ if we can make $f(x)$ arbitrarily large (and positive) by taking x sufficiently close to a (on either side of a) without letting $x = a$.

There is a similar definition for $\lim_{x\to a} f(x) = -\infty$ except we make $f(x)$ arbitrarily large and negative.

Relationship between the limit and one-sided limits

$\lim_{x\to a} f(x) = L \Rightarrow \lim_{x\to a^+} f(x) = \lim_{x\to a^-} f(x) = L$ $\qquad$ $\lim_{x\to a^+} f(x) = \lim_{x\to a^-} f(x) = L \Rightarrow \lim_{x\to a} f(x) = L$

$\lim_{x\to a^+} f(x) \neq \lim_{x\to a^-} f(x) \Rightarrow \lim_{x\to a} f(x)$ Does Not Exist

Properties

Assume $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist and c is any number then,

1. $\lim_{x\to a}\left[cf(x)\right] = c\lim_{x\to a} f(x)$
2. $\lim_{x\to a}\left[f(x) \pm g(x)\right] = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x)$
3. $\lim_{x\to a}\left[f(x)g(x)\right] = \lim_{x\to a} f(x)\ \lim_{x\to a} g(x)$
4. $\lim_{x\to a}\left[\frac{f(x)}{g(x)}\right] = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$ provided $\lim_{x\to a} g(x) \neq 0$
5. $\lim_{x\to a}\left[f(x)\right]^n = \left[\lim_{x\to a} f(x)\right]^n$
6. $\lim_{x\to a}\left[\sqrt[n]{f(x)}\right] = \sqrt[n]{\lim_{x\to a} f(x)}$

Basic Limit Evaluations at $\pm\infty$

T-test

One-Sample

Tests whether the mean of a normally distributed population is different from a specified value

Null Hypothesis (H_0): states that the population mean is equal to some value (μ_0)
Alternative Hypothesis (H_a): states that the mean does not equal/is greater than/is less than μ_0
t-statistic: standardizes the difference between $\bar{x}$ and μ_0

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Degrees of freedom (df) = *n-1*

Read the table of t-distribution critical values for the p-value (probability that the sample mean was obtained by chance given μ_0 is the population mean) using the calculated t-statistic and degrees of freedom.

- H_a: $\mu>\mu_0$ → the t-statistic is likely positive; read table as given
- H_a: $\mu<\mu_0$ → the t-statistic is likely negative; the t-distribution is symmetrical so read the probability as if the t-statistic were positive
- Note: if the t-statistic is of the 'wrong' sign, the p-value is 1 minus the *p* given in the chart
- H_a: $\mu\neq\mu_0$ → read the p-value as if the t-statistic were positive and double it (to consider both less than and greater than)

If the p-value is less than the predetermined value for significance (called α and is usually 0.05), reject the null hypothesis and accept the alternative hypothesis.

Example:
You are experiencing hair loss and skin discoloration and think it might be because of selenium toxicity. You decide to measure the selenium levels in your tap water once a day for one week. Your results are given below. The EPA maximum contaminant level for safe drinking water is 0.05 mg/L. Does the selenium level in your tap water exceed the legal limit (assume α=0.05)?

Day	Selenium mg/L
1	0.051
2	0.0505
3	0.049
4	0.0516
5	0.052
6	0.0508
7	0.0506

H_0: μ=0.05; H_a: μ>0.05
Calculate the mean and standard deviation of your sample:

$$\bar{x} = 0.0508$$

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{(0.051-0.0508)^2 + (0.0505-0.0508)^2 + etc\ldots}{6} = 9.15\times10^{-7}$$

$$s = \sqrt{s^2} = 9.56\times10^{-4}$$

The t-statistic is: $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.0508-0.05}{\frac{9.56\times10^{-4}}{\sqrt{7}}} = 2.17$ and the degrees of freedom are *n-1* = 7-1 = 6

Looking at the t-distribution of critical values table, 2.17 with 6 degrees of freedom is between *p*=0.05 and *p*=0.025. This means that the p-value is less than 0.05, so you can reject H_0 and conclude that the selenium level in your tap water exceeds the legal limit.

T-test

Two-Sample

Tests whether the means of two populations are significantly different from one another

Paired

Each value of one group corresponds directly to a value in the other group; ie: before and after values after drug treatment for each individual patient

Subtract the two values for each individual to get one set of values (the differences) and use μ_0 = 0 to perform a one-sample t-test

Unpaired

The two populations are independent

H_0: states that the means of the two populations are equal ($\mu_1=\mu_2$)

H_a: states that the means of the two populations are unequal or one is greater than the other ($\mu_1\neq\mu_2$, $\mu_1>\mu_2$, $\mu_1<\mu_2$)

Statistics Cheat Sheet

Population

The entire group one desires information about

Sample

A subset of the population taken because the entire population is usually too large to analyze
Its characteristics are taken to be representative of the population

Mean

Also called the arithmetic mean or average
The sum of all the values in the sample divided by the number of values in the sample/population
μ is the mean of the population; $\bar{x}$ is the mean of the sample

Median

The value separating the higher half of a sample/population from the lower half
Found by arranging all the values from lowest to highest and taking the middle one (or the mean of the middle two if there are an even number of values)

Variance

Measures dispersion around the mean
Determined by averaging the squared differences of all the values from the mean
Variance of a population is σ^2

$$\sigma^2 = \frac{\sum (x-\mu)^2}{n}$$

Can be calculated by subtracting the square of the mean from the average of the squared scores:

$$\sigma^2 = \frac{\sum x^2}{n} - \mu^2$$

Variance of a sample is s^2; note the *n-1*

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

Can be calculated by:

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

Standard Deviation

Square root of the variance
Also measures dispersion around the mean but in the same units as the values (instead of square units with variance)
σ is the standard deviation of the population and s is the standard deviation of the sample

Standard Error

An estimate of the standard deviation of the sampling distribution—the set of all samples of size n that can be taken from a population
Reflects the extent to which a statistic changes from sample to sample

For a mean, $\frac{s}{\sqrt{n}}$

For the difference between two means,

Assuming equal variances $\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$; unequal variances $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$