# Review of Linear Algebra

This review covers linear algebra material you have learned in a prerequisite course.

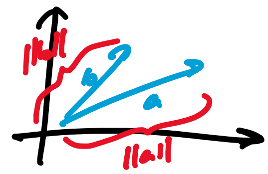
It is intended as a quick refresher of the most relevant parts of linear algebra for the early weeks of MTH 4224 Intro to ML.

More ideas from linear algebra, multivariate calculus, probability, and statistics will be taught in class as needed.

Vectors

Let $a, b \in \mathbb{R}^d$, $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{bmatrix}$

2D vectors




$$a \cdot b = a^T b = a_1 b_1 + a_2 b_2 + \dots + a_d b_d = \sum_{i=1}^{d} a_i b_i$$

is the <u>dot product</u> of a and b

$$\|a\| = \sqrt{a^T a} = \sqrt{a_1^2 + a_2^2 + \dots + a_d^2} = \left( \sum_{i=1}^{d} |a_i|^2 \right)^{1/2}$$ is the

<u>Euclidean length</u> of a (or the <u>$L^2$ norm</u> of a)

<u>Note</u>: $\|a\|^2 = a^T a$ ← "sum of squares" $a_1^2 + a_2^2 + \dots + a_d^2$

$$\|a\|_p = \left( \sum_{i=1}^{d} |a_i|^p \right)^{1/p}$$ is the <u>$L^p$ norm</u> of a (for any $p \neq 0$)

A <u>unit vector</u> has norm $= 1$. A unit vector in the direction of $a = \frac{a}{\|a\|}$

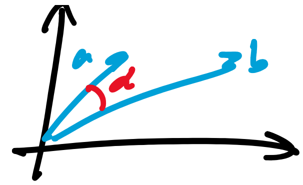
The <u>Euclidean distance</u> between a and b is $\|a - b\|$

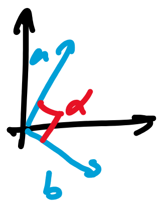
The <u>angle</u> between a and b satisfies $\cos \alpha = \frac{a^T b}{\|a\| \cdot \|b\|}$

... if $a^T b = 0$

a and b are <u>orthogonal</u> if $a^T b = 0$

(i.e. $\alpha = \frac{\pi}{2}$)



# Matrices

$A \in \mathbb{R}^{n \times d}$ is a <u>matrix</u>

row vectors $\in \mathbb{R}^d$ column vectors $\in \mathbb{R}^n$

$$A = \begin{array}{cccc|c} & & & & \\ \end{array}\begin{array}{c} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{bmatrix} \\ \begin{array}{cccc} \uparrow & \uparrow & & \uparrow \\ \text{col } 1 & \text{col } 2 & & \text{col } d \end{array} \end{array} \begin{array}{l} \leftarrow \text{row } 1 \\ \leftarrow \text{row } 2 \\ \vdots \\ \leftarrow \text{row } n \\ \\ \\ \end{array} = \begin{bmatrix} \text{---} & a_1 & \text{---} \\ \text{---} & a_2 & \text{---} \\ & \vdots & \\ \text{---} & a_n & \text{---} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A_1 & A_2 & \cdots & A_d \\ | & | & & | \end{bmatrix}$$

$$I_n = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix}_{n \times n}$$ is the <u>identity matrix</u>

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1d} \\ & u_{22} & \cdots & u_{2d} \\ 0 & & \ddots & u_{nd} \end{bmatrix} \quad \& \quad L = \begin{bmatrix} l_{11} & & 0 \\ l_{21} & l_{22} & \\ \vdots & \vdots & \ddots \\ l_{n1} & l_{n2} & \cdots & l_{nd} \end{bmatrix}$$ are <u>upper triangular</u> and <u>lower triangular</u> matrices, respectively

$$D = \begin{bmatrix} d_{11} & & 0 \\ & d_{22} & \\ 0 & & \ddots & \\ & & & d_{nn} \end{bmatrix}$$ is a <u>diagonal matrix</u>

# Matrix Multiplication

it is not elementwise multiplication

must be the same to multiply matrices

Let $A \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{d \times m}$

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{bmatrix}_{n \times d} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{d1} & b_{d2} & \cdots & b_{dm} \end{bmatrix}_{d \times m} = \begin{bmatrix} \sum_{i=1}^{d} a_{1i} b_{i1} & \sum_{i=1}^{d} a_{1i} b_{i2} & \cdots & \sum_{i=1}^{d} a_{1i} b_{im} \\ \sum_i a_{2i} b_{i1} & \sum_i a_{2i} b_{i2} & \cdots & \sum_i a_{2i} b_{im} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i a_{ni} b_{i1} & \sum_i a_{ni} b_{i2} & \cdots & \sum_i a_{ni} b_{im} \end{bmatrix}_{n \times m}$$

$a_2$ $\quad B_1$

$$= \begin{bmatrix} a_1 \cdot B_1 & a_1 \cdot B_2 & \cdots & a_1 \cdot B_m \\ a_2 \cdot B_1 & a_2 \cdot B_2 & \cdots & a_2 \cdot B_m \\ \vdots & \vdots & \ddots & \vdots \\ a_n \cdot B_1 & a_n \cdot B_2 & \cdots & a_n \cdot B_m \end{bmatrix}$$

all of these properties assume the matrices are shaped such that these operations are defined

Properties: $AB \neq BA$ in general — matrix multiplication is not commutative

distributive properties: $A(B+C) = AB + AC$ & $(A+B)C = AC + BC$

$$c(AB) = (cA)B = A(cB) = (AB)c$$

constant

associative property $\quad (AB)C = A(BC)$

identity matrix multiplication

$$AI = A$$

$$IA = A$$

Computational Notes:

As of 2020, the fastest algorithm is $O(n^{2.3728596})$ for $n \times n$ matrices

Special types of matrices can be faster
(e.g symmetric, banded, triangular, sparse, diagonal matrices) mostly 0's

We can safely rely on tools like numPy, tensorflow, CUDA, MATLAB, etc. to take care of this efficiently

# Matrix Inverses

Assume $A$ is a square matrix, i.e. $A \in \mathbb{R}^{n \times n}$

A matrix $A^{-1}$ such that $AA^{-1} = A^{-1}A = I_n$ is the <u>matrix inverse</u> of $A$

If $A$ has an inverse, it is <u>invertible</u>, otherwise it is <u>singular</u>

↳ $A$ being invertible is central to elementary linear algebra (see the invertible matrix theorem)

<u>properties</u>: $(kA)^{-1} = \frac{1}{k}A^{-1}$

$$(A^T)^{-1} = (A^{-1})^T$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

## Computational Note

Matrix inversion is $O(n^{2.373}) \rightarrow$ tools will do it efficiently

# Linear Systems

Let $A \in \mathbb{R}^{n \times d}$, $x, b \in \mathbb{R}^d$, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{bmatrix}$

$$\begin{matrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1d}x_d = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2d}x_d = b_2 \\ \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nd}x_d = b_d \end{matrix} \longrightarrow \begin{bmatrix} a_1 \cdot x \\ a_2 \cdot x \\ \vdots \\ a_n \cdot x \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{bmatrix}$$

$$Ax = b$$

system of linear equations can be written compactly as a matrix equation

If $A$ is invertible, we can solve for the vector $x$:

$$(A^{-1}A)x = A^{-1}b$$

$$Ix = A^{-1}b$$

$$x = A^{-1}b$$

If $A$ is singular, the system does not have a unique solution

Matrix Transpose

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{bmatrix} = \begin{bmatrix} - & a_1 & - \\ - & a_2 & - \\ & \vdots & \\ - & a_n & - \end{bmatrix}$$

The <u>transpose</u> of $A$ is

swap rows and columns

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1d} & a_{2d} & \cdots & a_{nd} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a_1^T & a_2^T & \cdots & a_n^T \\ | & | & & | \end{bmatrix}$$

<u>properties</u> $(AB)^T = B^T A^T$

If $A^T = A$, then $A$ is a <u>symmetric matrix</u>

If inverse $A^{-1}$ exists, it is symmetric **if and only if $A$ is symmetric**.

# Eigenvalues and Eigenvectors

Let $A \in \mathbb{R}^{n \times n}$, $u \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$

If $Au = \lambda u$, $u$ is an **eigenvector** of $A$

$\lambda$ is an **eigenvalue** of $A$

a vector that points in the same direction when multiplied by A

## How do we find eigenvectors?

$$Au = \lambda u \Rightarrow Au - \lambda u = 0$$

$$(A - \lambda I)u = 0$$

By the invertible matrix theorem, this system has a nonzero solution if and only if

computed at $O(n^{2.376})$

$$\det(A - \lambda I) = 0$$

fundamental theorem of algebra

$$(\lambda - \lambda_1) \cdots (\lambda - \lambda_n) = 0$$

where $\lambda_1, \ldots, \lambda_n$ are solutions (i.e. the eigenvalues)

Matrix Calculus

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function, $x \in \mathbb{R}^d$, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$

(vector x, scaler f(x))

The gradient of f is $\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{bmatrix} = \frac{\partial f}{\partial x}$

Identities: Assume $f : \mathbb{R}^d \to \mathbb{R}$, $g : \mathbb{R}^d \to \mathbb{R}$, $a, b \in \mathbb{R}$

$\nabla(af(x) + bg(x)) = a\nabla f(x) + b\nabla g(x)$ gradiant is a linear operator

$\nabla(f(x)g(x)) = f(x)\nabla g(x) + g(x)\nabla f(x)$ product rule