W203 Linear Algebra / Matrix Review

- Linear algebra concerns itself with "mathematical structures closed under the operations of addition and scalar multiplication. . ." – Merriam Webster Dictionary

- A function $f: \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear if for any two "vectors" (inputs) $u, v \in \mathbb{R}^n$ and any scalar $c \in \mathbb{R}$, we have the following:

  - $f(u+v) = f(u) + f(v)$ additivity
  - $f(cu) = cf(u)$ scalar multpiplication

line through origin

$\boxed{\text{Ex}}$ Let $f(x) = 5x$ $\quad f: \mathbb{R} \to \mathbb{R}$

- $f(u+v) = 5(u+v) = 5u + 5v = f(u) + f(v)$ additivity ✓
- $f(cu) = 5(cu) = c(5u) = cf(u)$ ✓ scalar multiplication

$\boxed{\text{Non-ex}}$ Let $f(x) = x^2$

$$f(u+v) = u^2 + 2uv + v^2 \neq u^2 + v^2 = f(u) + f(v)$$
$$f(cu) = (cu)^2 = c^2u^2 \neq cu^2 = cf(u)$$

Some special examples:

derivatives: Let "vectors" / inputs be functions $u(x), v(x)$ and f be the derivative operator d/dx

$$f(u+v) = \frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx} = f(u) + f(v)$$
$$f(cu(x)) = d/dx\,(cu(x)) = c\frac{du}{dx}$$

summations Σ

integrals $\int$

Momentary Context Switch

Putting the linear function story aside for a moment, let's look at a popular object in Data Science & Machine Learning: the matrix

- A matrix is a rectangular array of "elements" arranged in rows & columns, it is a way of organizing information (data structure)

- As far as we are concerned, a matrix $A \in \mathbb{R}^{m \times n}$ is a rectangular array of $\underbrace{\text{real numbers}}_{\text{"elements"}}$ with m rows & n columns

elements/entries of a matrix are denoted by
a_{ij}, where $i = 1, 2, \ldots m$ denotes rows
$j = 1, 2, \ldots n$ denotes columns

Ex $A \in \mathbb{R}^{3 \times 2}$ $\begin{bmatrix} 1 & 8 \\ 2 & 9 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}$

4

Some more vocab:
Dimension/Order of a matrix is "num rows by num columns" $(m \times n)$

Row vector $\underline{v}$ is a matrix of order $1 \times n$
Column vector $\underline{w}$ is a matrix of order $m \times 1$
ex $v = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ $w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

Operations with Matrices

Transpose

"flip" rows & columns

$a_{ij} \rightarrow a_{ji}$

$(A^T)_{ij} = A_{ji}$ or sometimes written as $\underline{A}'$

Notes: If $\underline{A} \in \mathbb{R}^{m \times n}$, $\underline{A}^T \in \mathbb{R}^{n \times m}$

- From now on assume all vectors $\underline{v}$ we talk about are column vectors i.e $\underline{v} \in \mathbb{R}^{m \times 1}$ since any column vector can be attained as $\underline{v}^T$

Addition / Subtraction

* Matrices need to be of same dimension

$\underline{A} \pm \underline{B} = \underline{C}$ s.t. $c_{ij} = a_{ij} \pm b_{ij}$, $\underline{A}, \underline{B}, \underline{C} \in \mathbb{R}^{m \times n}$

ex $\begin{bmatrix} 1 & 3 \\ 100 & 70 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 101 & 75 \end{bmatrix}$

$\underline{A} \in \mathbb{R}^{2 \times 2}$ $\quad$ $\underline{B} \in \mathbb{R}^{2 \times 2}$ $\quad$ $\underline{C} \in \mathbb{R}^{2 \times 2}$

Multiplication

*number of columns of $\underline{A}$ = number of rows of $\underline{B}$

If $\underline{A} \in \mathbb{R}^{m \times n}$, $\underline{B} \in \mathbb{R}^{n \times p}$ then $\underline{C} = \underline{A}\underline{B} \in \mathbb{R}^{m \times p}$

The formula is given as: $c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$

→ ij^{th} element found by multiplying elements of the i^{th} row of A with j^{th} column of B and summing

Ex $\underline{C} = \underline{A}\,\underline{D} = \begin{bmatrix} 6 & 8 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 3 & -8 & 1 \\ 9 & 2 & 5 \end{bmatrix}$

$$= \begin{bmatrix} 6*3+8*9 & 6*(-8)+8*2 & 6*1+8*5 \\ (-2)*3+4*9 & (-2)*(-8)+4*2 & (-2)*1+4*5 \end{bmatrix}$$

$$= \begin{bmatrix} 90 & -32 & 46 \\ 30 & 24 & 18 \end{bmatrix}$$

Symbolically

$$C = AB = \begin{bmatrix} - a_1^T - \\ - a_2^T - \\ \vdots \\ - a_m^T - \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ \vdots & & & \vdots \\ a_m^T b_1 & \cdots & \cdots & a_m^T b_p \end{bmatrix}$$

Note: The "naturally expected" product of $c_{ij} = a_{ij} * b_{ij}$ is known as the Kroenecker Product & is usually NOT the product we mean when we discuss matrix multiplication

Matrix Inverse

* only properly defined for square matrices (number rows = number columns)

defined as matrix $\underline{A}^{-1}$ s.t. $\underline{A}^{-1}\underline{A} = I$ for $A \in \mathbb{R}^{n \times n}$

where I is the identity matrix : $\underbrace{\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}}_{n} \Big\} n$

$I_{ij} = 1$ if $i = j$
0 else

* closest parallel to number 1

Note: Even if square $\not\Rightarrow$ invertible / A^{-1} exists, e.g matrix of all 0's

Why a "weird" multiplication & division/inverse?

$\boxed{\text{Key Result: Linear functions can be represented as Matrices}}$ ← what we set aside earlier

So for fxn f, we have corresponding $\underline{A}_f$ s.t. $f(x) \Leftrightarrow \underline{A}_f x$

- Matrix multiplication as defined is paralleled by function composition $g \circ f = \underbrace{g(f(x)) \Longleftrightarrow \underline{B}_g A_f \underline{x}}_{\text{order of multiplication/composition matters!}}$

- $f^{-1} \Longleftrightarrow A_f^{-1}$, not all functions/square matrices are invertible

* Practical Result: Computers can't symbollicaly do derivatives, integrals, etc. but they can store & operate matrices!!!

Same idea applied along stats, machine learning, etc.

Some examples can look up:

- Principle Component Analysis (PCA)
- Matrix Decomposition, Recommender Systems
- Image Proccessing with Gaussian Blur

Key Properties of Operations

Scale
- For $c \in \mathbb{R}$, $\underline{A} \in \mathbb{R}^{m \times n}$ $\quad c\underline{A} = c * a_{ij} \ \forall i,j$
- $\underline{A}\underline{\underline{I}} = \underline{A} = \underline{\underline{I}}$ for $\underline{A}, \underline{\underline{I}} \in \mathbb{R}^{m \times m}$, $\underline{\underline{I}}$ identity matrix

Transpose
- $(\underline{A}^T)^T = \underline{A}$
- $(\underline{A}\underline{B})^T = \underline{B}^T \underline{A}^T$
- $(\underline{A} \pm \underline{B})^T = \underline{A}^T \pm \underline{B}^T$
- $(c\underline{A})^T = c\underline{A}^T$, $c \in \mathbb{R}$
- $A^T A$ will always be square

Inverse
- $(\underline{A}^{-1})^{-1} = \underline{A}$
- $(\underline{A}\underline{B})^{-1} = \underline{B}^{-1}\underline{A}^{-1}$
- $(\underline{A}^{-1})^T = (\underline{A}^T)^{-1}$

Addition
- $\underline{A} \pm \underline{B} = \underline{B} \pm \underline{A}$
- $(\underline{A} \pm \underline{B}) \pm \underline{C} = \underline{A} \pm (\underline{B} \pm \underline{C})$

Multiplication:
- $\underline{A}(\underline{B}+\underline{C}) = \underline{A}\underline{B} + \underline{A}\underline{C}$
- $(\underline{A}+\underline{B})\underline{C} = \underline{A}\underline{C} + \underline{B}\underline{C}$
- $\underline{A}(\underline{B}\underline{C}) = (\underline{A}\underline{B})\underline{C}$
- $(\underline{A}\underline{B})^T = B^T A^T$
- $(ABC)' = C^T B^T A^T$

Matrices as systems of Linear Equations

*Historical / original application of matrices as compact representation

Consider: $1x_1 + 2x_2 = 5$
$3x_1 + 9x_2 = 21$

From algebra, we know how to solve via substitution, elimination, subtraction to find x_1, x_2

Solution can be unique, no solution, or infinitely many

In matrix notation, we can write this system as

$$\underline{A}\,\underline{x} = \underline{b}$$

with $\underline{A} = \begin{bmatrix} 1 & 2 \\ 3 & 9 \end{bmatrix}$, $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $b = \begin{bmatrix} 5 \\ 21 \end{bmatrix}$

Solving for $\underline{x}$, we have $\underline{A}^{-1}\underline{A}\,\underline{x} = \underline{A}^{-1}\underline{b}$

$$\Rightarrow \underline{I}\,\underline{x} = \underline{A}^{-1}\underline{b}$$

$$\Rightarrow \underline{x} = \underline{A}^{-1}\underline{b}$$

claim $A^{-1} = \begin{bmatrix} 3 & -2/3 \\ -1 & 1/3 \end{bmatrix}$ can check by computing $A^{-1}b$ and plugging in ~~x~~ found

Can do this by hand via Guass-Jordan Elimination (won't cover this)

Vector/Matrix Calculus

Recall partial derivative $\frac{\partial}{\partial x}$ of $f(x,y)$ is like our usual scalar derivative in that we "treat the other variable(s) as constants"

e.g. $f(x,y) = 3x^2y$

$$\frac{\partial f}{\partial x} = 6xy$$

$$\frac{\partial f}{\partial y} = 3x^2$$

We can organize as a vector function known as gradient denoted as "∇"

$$\nabla f(x,y) = \begin{bmatrix} \partial f/\partial x \\ \partial f/\partial y \end{bmatrix} = \begin{bmatrix} 6xy \\ 3x^2 \end{bmatrix}$$

We can jump from vectors to matrices when we have to take derivatives of multiple functions

consider again $f(x,y) = 6xy$

and now $g(x,y) = 2x + y^8$

If we take ∇ of f and g, get a matrix known as the Jacobian

$$J = \begin{bmatrix} \nabla f(x,y) \\ \nabla g(x,y) \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix} = \begin{bmatrix} 6xy & 3x^2 \\ 2 & 8y^7 \end{bmatrix}$$

Note: Some texts use J^T as their Jacobian

Compacting notation further we can write

$f(x, y) = f(\underline{u})$ $\quad \underline{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

So now we can write multiple scalar-valued functions f_i as function vector $\underline{f}$

Let $\underline{y} = \underline{f}(\underline{x})$

$$\rightarrow y_1 = f_1(\underline{x}) \qquad \text{eg } f_1 = f$$

$$y_2 = f_2(\underline{x}) \qquad f_2 = g$$

from before

$$\vdots$$

$$y_m = f_m(\underline{x})$$

So we can generally write

$$\frac{\partial \underline{y}}{\partial \underline{x}} = \begin{bmatrix} \nabla f_1(\underline{x}) \\ \vdots \\ \nabla f_m(\underline{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \underline{x}} f_1(\underline{x}) \\ \vdots \\ \frac{\partial}{\partial \underline{x}} f_m(\underline{x}) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(\underline{x}) & \cdots & \frac{\partial}{\partial x_n} f_1(\underline{x}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} f_m(\underline{x}) & & \frac{\partial}{\partial x_n} f_m(\underline{x}) \end{bmatrix}$$

There are actually many other definitions in terms of defining

$\frac{\partial A}{\partial x}$ matrix wrt scalar

$\frac{\partial y}{\partial A}$ scalar wrt matrix ...

but we leave to curious minds to explore beyond ...