Lecture 2: Linear algebra

STATS 101: Foundations of Statistics

Linh Tran
ThetaHat.AI@gmail.com

November 21, 2019

Announcements

- 30 students submitted homework
- Next assignment will be posted tonight (due 12/4 @ 11:59pm)
- Re: prizes for top 3 students
  - Amazon gift cards
  - Extra points are awarded for:
    1. Class participation (e.g. asking/answering questions, etc)
    2. Catching & correcting errata/typos
    3. Answering questions / participating in discussions on Piazza
  - Blinded top 3 scores will be posted to course website
- No class next week (Happy Thanksgiving!)

Outline

All things linear algebra

- Basic concepts
- Matrix multiplication
- Operations and Properties
- Matrix Calculus

Basic concepts

Consider the following equations:

$$
\begin{aligned}
4x_1 - 5x_2 &= -13 \qquad (1) \\
-2x_1 + 3x_2 &= 9 \qquad (2)
\end{aligned}
$$

Let's solve for x_1 and x_2.

Basic concepts

Consider the following equations:

$$\begin{aligned} 4x_1 - 5x_2 &= -13 \qquad (1) \\ -2x_1 + 3x_2 &= 9 \qquad (2) \end{aligned}$$

Let's solve for x_1 and x_2.

We can write this system of equations more compactly in matrix notation, e.g.

$$\mathbf{A}\mathbf{x} = \mathbf{b} \qquad (3)$$

where $\mathbf{A} = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$

Basic concepts

Some basic notation:

- We denote a matrix with m rows and n columns as $\mathbf{A} \in \mathbb{R}^{m \times n}$, where each entry in the matrix is a real number.
- We denote a vector with n entries as $\mathbf{x} \in \mathbb{R}^n$.
  - By convention, we typically think of a vector as a 1 column matrix.
- We denote the i^{th} element of a vector $\mathbf{x}$ as x_i, e.g.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{4}$$

Basic concepts

Some basic notation:

- We denote each entry in a matrix $\mathbf{A}$ by a_{ij}, corresponding to the i^{th} row and j^{th} column, e.g.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{5}$$

- We denote the *transpose* of a matrix as $\mathbf{A}^\top$, e.g.

$$\mathbf{A}^\top = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \tag{6}$$

Basic concepts

Some basic notation:

- We denote the j^{th} column of $\mathbf{A}$ by $\mathbf{a}_j$ or $\mathbf{A}_{\cdot j}$, e.g.

$$\mathbf{A} = \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} \tag{7}$$

- We denote the i^{th} row of $\mathbf{A}$ by $\mathbf{a}_i^\top$ or $\mathbf{A}_{i\cdot}$.

$$\mathbf{A} = \begin{bmatrix} — & \mathbf{a}_1^\top & — \\ — & \mathbf{a}_2^\top & — \\ & \vdots & \\ — & \mathbf{a}_m^\top & — \end{bmatrix} \tag{8}$$

n.b. This isn't universal, though should be clear from its presentation and use.

Matrix multiplication

Given two matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$, we can multiply them by

$$\mathbf{C} = \mathbf{AB} \in \mathbb{R}^{m \times p} : \mathbf{C}_{ij} = \sum_{k=1}^{n} \mathbf{A}_{ik} \mathbf{B}_{kj} \tag{9}$$

n.b. The dimensions have to be compatible for matrix multiplication to be valid (e.g. the number of columns in $\mathbf{A}$ must be equal to the number of rows in $\mathbf{B}$).

Matrix multiplication

Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the quantity $\mathbf{x}^\top \mathbf{y} \in \mathbb{R}$ (aka *dot product* or *inner product*) is a scalar given by

$$\mathbf{x}^\top \mathbf{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i \tag{10}$$

Note: For vectors, we always have that $\mathbf{x}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{x}$. This is not generally true for matrices.

Matrix multiplication

Given $\mathbf{x} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{R}^n$, the quantity $\mathbf{x}^\top \mathbf{y} \in \mathbb{R}^{m \times n}$ (aka *outer product*) is a matrix given by

$$\mathbf{x}\mathbf{y}^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix} \tag{11}$$

Matrix multiplication

Example: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix such that all columns are equal to some vector $\mathbf{x} \in \mathbb{R}^m$. Using outer products, we can represent $\mathbf{A}$ compactly as

$$
\begin{aligned}
\mathbf{A} = \begin{bmatrix} | & | & & | \\ \mathbf{x} & \mathbf{x} & \cdots & \mathbf{x} \\ | & | & & | \end{bmatrix} &= \begin{bmatrix} x_1 & x_1 & \cdots & x_1 \\ x_2 & x_2 & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_m & \cdots & x_m \end{bmatrix} && (12) \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} && (13) \\
&= \mathbf{x}\mathbf{1}^\top && (14)
\end{aligned}
$$

Matrix-vector products

Given $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^n$, their product is a vector $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$.

Matrix-vector products

Given $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^n$, their product is a vector $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$.

There are two ways of interpreting this:

$$
\begin{aligned}
\mathbf{y} = \mathbf{A}\mathbf{x} &= \begin{bmatrix} — & \mathbf{a}_1^\top & — \\ — & \mathbf{a}_2^\top & — \\ & \vdots & \\ — & \mathbf{a}_m^\top & — \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{x} \\ \mathbf{a}_2^\top \mathbf{x} \\ \vdots \\ \mathbf{a}_m^\top \mathbf{x} \end{bmatrix} \qquad (15) \\
&= \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \qquad (16) \\
&= \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_n x_n \qquad (17)
\end{aligned}
$$

Matrix-vector products

Example:

Define $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} -3 \\ -2 \\ -1 \end{bmatrix}.$

Calculate $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Matrix-matrix products

Given $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}$, their product is a matrix $\mathbf{C} = \mathbf{AB} \in \mathbb{R}^{m \times p}$.

Matrix-matrix products

Given $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}$, their product is a matrix $\mathbf{C} = \mathbf{A}\mathbf{B} \in \mathbb{R}^{m \times p}$.

Similar to before, we can think of this in two ways:

Interpretation # 1

$$\begin{aligned}
\mathbf{C} = \mathbf{A}\mathbf{B} &= \begin{bmatrix} — & \mathbf{a}_1^\top & — \\ — & \mathbf{a}_2^\top & — \\ & \vdots & \\ — & \mathbf{a}_m^\top & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \\ | & | & & | \end{bmatrix} && (18) \\
&= \begin{bmatrix} \mathbf{a}_1^\top \mathbf{b}_1 & \mathbf{a}_1^\top \mathbf{b}_2 & \cdots \mathbf{a}_1^\top \mathbf{b}_p \\ \mathbf{a}_2^\top \mathbf{b}_1 & \mathbf{a}_2^\top \mathbf{b}_2 & \cdots \mathbf{a}_2^\top \mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m^\top \mathbf{b}_1 & \mathbf{a}_m^\top \mathbf{b}_2 & \cdots \mathbf{a}_m^\top \mathbf{b}_p \end{bmatrix} && (19)
\end{aligned}$$

Matrix-matrix products

Interpretation # 2

$$\mathbf{C} = \mathbf{A}\mathbf{B} = \mathbf{A}\begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \\ | & | & & | \end{bmatrix} \tag{20}$$

$$= \begin{bmatrix} | & | & & | \\ \mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 & \cdots & \mathbf{A}\mathbf{b}_p \\ | & | & & | \end{bmatrix} \tag{21}$$

$$= \begin{bmatrix} — & \mathbf{a}_1^\top & — \\ — & \mathbf{a}_2^\top & — \\ & \vdots & \\ — & \mathbf{a}_m^\top & — \end{bmatrix}\mathbf{B} = \begin{bmatrix} — & \mathbf{a}_1^\top\mathbf{B} & — \\ — & \mathbf{a}_2^\top\mathbf{B} & — \\ & \vdots & \\ — & \mathbf{a}_m^\top\mathbf{B} & — \end{bmatrix} \tag{22}$$

Matrix multiplication properties

- Associative: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- Distributive: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- Not commutative: $\mathbf{AB} \neq \mathbf{BA}$

Matrix multiplication properties

Demonstrating *associativity*:

We just need to show that $((\mathbf{AB})\mathbf{C})_{ij} = (\mathbf{A}(\mathbf{BC}))_{ij}$:

$$\begin{aligned}
((\mathbf{AB})\mathbf{C})_{ij} &= \sum_{k=1}^{p}(\mathbf{AB})_{ik}\mathbf{C}_{kj} = \sum_{k=1}^{p}\left(\sum_{l=1}^{n}\mathbf{A}_{il}\mathbf{B}_{lk}\right)\mathbf{C}_{kj} && (23)\\
&= \sum_{k=1}^{p}\left(\sum_{l=1}^{n}\mathbf{A}_{il}\mathbf{B}_{lk}\mathbf{C}_{kj}\right) = \sum_{l=1}^{n}\left(\sum_{k=1}^{p}\mathbf{A}_{il}\mathbf{B}_{lk}\mathbf{C}_{kj}\right) && (24)\\
&= \sum_{l=1}^{n}\mathbf{A}_{il}\left(\sum_{k=1}^{p}\mathbf{B}_{lk}\mathbf{C}_{kj}\right) = \sum_{l=1}^{n}\mathbf{A}_{il}(\mathbf{BC})_{lj} && (25)\\
&= (\mathbf{A}(\mathbf{BC}))_{ij} && (26)
\end{aligned}$$

The identity matrix:

The *identity matrix*, denoted $\mathbf{I} \in \mathbb{R}^{n \times n}$ is a square matrix with 1's in the diagonal and 0's everywhere else, i.e.

$$\mathbf{I}_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{27}$$

Operations & properties

The identity matrix:

The *identity matrix*, denoted $\mathbf{I} \in \mathbb{R}^{n \times n}$ is a square matrix with 1's in the diagonal and 0's everywhere else, i.e.

$$\mathbf{I}_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{27}$$

It has the property

$$\mathbf{AI} = \mathbf{A} = \mathbf{IA} \; \forall \mathbf{A} \in \mathbb{R}^{m \times n} \tag{28}$$

n.b. The dimensionality of $\mathbf{I}$ is typically inferred (e.g. $n \times n$ vs $m \times m$)

The diagonal matrix: The *diagonal matrix*, denoted $\mathbf{D} = diag(d_1, d_2, \ldots, d_n)$ is a matrix where all non-diagonal elements are 0, i.e.

$$\mathbf{D}_{ij} = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases} \tag{29}$$

Clearly, $\mathbf{I} = diag(1, 1, ..., 1)$.

The transpose

The *transpose* of a matrix results from "*flipping*" the rows and columns, i.e.

$$(\mathbf{A}^\top)_{ij} = \mathbf{A}_{ji} \tag{30}$$

Consequently, for $\mathbf{A} \in \mathbb{R}^{m \times n}$ we have that $\mathbf{A}^\top \in \mathbb{R}^{n \times m}$.

Some properties:

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

Symmetry

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *symmetric* if $\mathbf{A} = \mathbf{A}^\top$.

It is *anti-symmetric* if $\mathbf{A} = -\mathbf{A}^\top$.

Symmetry

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *symmetric* if $\mathbf{A} = \mathbf{A}^\top$.

It is *anti-symmetric* if $\mathbf{A} = -\mathbf{A}^\top$.

It is easy to show that $\mathbf{A} + \mathbf{A}^\top$ is symmetric and $\mathbf{A} - \mathbf{A}^\top$ is anti-symmetric. Consequently, we have that

$$\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^\top) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^\top) \tag{31}$$

Symmetry

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *symmetric* if $\mathbf{A} = \mathbf{A}^\top$.

It is *anti-symmetric* if $\mathbf{A} = -\mathbf{A}^\top$.

It is easy to show that $\mathbf{A} + \mathbf{A}^\top$ is symmetric and $\mathbf{A} - \mathbf{A}^\top$ is anti-symmetric. Consequently, we have that

$$\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^\top) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^\top) \tag{31}$$

Symmetric matrices tend to be denoted as $\mathbf{A} \in \mathbb{S}^n$.

Trace

The *trace* of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, denoted $tr(\mathbf{A})$ or $tr\mathbf{A}$ is the sum of the diagonal elements, i.e.

$$tr\mathbf{A} = \sum_{i=1}^{n} \mathbf{A}_{ii} \tag{32}$$

The trace has the following properties:

- For $\mathbf{A} \in \mathbb{R}^{n \times n}$, $tr\mathbf{A} = tr\mathbf{A}^\top$
- For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, $tr(\mathbf{A} + \mathbf{B}) = tr\mathbf{A} + tr\mathbf{B}$
- For $\mathbf{A} \in \mathbb{R}^{n \times n}, c \in \mathbb{R}$, $tr(c\mathbf{A}) = c\, tr\mathbf{A}$
- For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n} \ni \mathbf{AB} \in \mathbb{R}^{n \times n}$, $tr\mathbf{AB} = tr\mathbf{BA}$
- For $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}^{n \times n} \ni \mathbf{ABC} \in \mathbb{R}^{n \times n}$, $tr\mathbf{ABC} = tr\mathbf{BCA} = tr\mathbf{CAB}$, and so on for more matrices

Trace

Example: Proving that $tr\mathbf{AB} = tr\mathbf{BA}$

$$\begin{aligned}
tr\mathbf{AB} &= \sum_{i=1}^{m} (\mathbf{AB})_{ii} = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \mathbf{A}_{ij}\mathbf{B}_{ji} \right) && (33) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{A}_{ij}\mathbf{B}_{ji} = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{B}_{ji}\mathbf{A}_{ij} && (34) \\
&= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \mathbf{B}_{ji}\mathbf{A}_{ij} \right) = \sum_{j=1}^{n} (\mathbf{BA})_{jj} && (35) \\
&= tr\mathbf{BA} && (36)
\end{aligned}$$

References

Some textbooks on linear algebra:

- *Linear Algebra (Jim Hefferon)*
- *Introduction to Applied Linear Algebra (Boyd & Vandenberghe)*
- *Linear Algebra (Cherney, Denton et al.)*
- *Linear Algebra (Hoffman & Kunze)*
- *Fundamentals of Linear Algebra (Carrell)*
- *Linear Algebra (S. Friedberg A. Insel L. Spence)*