Lecture 6: Vectors and Matrices

Juan Dodyk

WashU

Plan

Vectors

1. Vectors in $\mathbb{R}^n$
2. Sum, inner product
3. Norm, triangle inequality, Cauchy-Schwarz

Matrices

1. Matrices in $\mathbb{R}^{n \times m}$
2. Sum, scalar product, matrix product
3. Vectors as matrices
4. Transpose


household
essentials
FROM
THE
PRODUCE

Vectors

$\mathbb{R}^{\boldsymbol{n}}$ is the set of **vectors** $(x_1, \ldots, x_n)$, which are lists of real numbers $x_1, \ldots, x_n$.

$\mathbb{R}^1$ is just $\mathbb{R}$.

$\mathbb{R}^2$ is the set of pairs (x_1, x_2). We can represent them as points (x, y) in the plane.

If we have n people and we have asked their age, we can represent that data as a vector in $\mathbb{R}^n$.

The zero vector $\mathbf{0} \in \mathbb{R}^n$ is defined as $(0, 0, \ldots, 0)$.

Some people use bold to emphasize that something is a vector, e.g., $\boldsymbol{x} \in \mathbb{R}^n$, $\mathbf{0} \in \mathbb{R}^n$. Others use a little arrow, e.g., $\vec{x} \in \mathbb{R}^n$, $\vec{0} \in \mathbb{R}^n$. I'll use bold here, but in general I don't think it's necessary.

Operations

Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we can take their sum or difference $\boldsymbol{x} \pm \boldsymbol{y}$, which is the vector

$$\boldsymbol{x} \pm \boldsymbol{y} = (x_1 \pm y_1, x_2 \pm y_2, \ldots, x_n \pm y_n).$$

Example.

- $(1, -1, 0) + (2, 3, -2) = (3, 2, -2)$.
- $(1, 0) + (0, 1, 2)$ is not defined.

Given $\boldsymbol{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$ we define the **scalar multiplication** as $ax \in \mathbb{R}^n$ given by

$$a\boldsymbol{x} = (ax_1, ax_2, \ldots, ax_n).$$

Example. $2 \cdot (1, 2, 3) = (2, 4, 6)$.

Inner Product

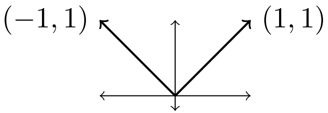
Given $x, y \in \mathbb{R}^n$ we define their **inner product** (or **dot product**, or **scalar product**) as the number

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Example. $(1, -1, 0) \cdot (2, 3, -2) = 2 - 3 + 0 = -1$.

If $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ we say that the vectors are **orthogonal**, or **perpendicular**.

Example. $(1, 1) \cdot (-1, 1) = 0$.



Euclidean Norm

Given $\boldsymbol{x} \in \mathbb{R}^n$ we define its (Euclidean) **norm** as

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}.$$

It measures the length of the line from the point $\mathbf{0}$ to the point $\boldsymbol{x}$.

Example. If $\boldsymbol{x} = (1,1)$, $\|\boldsymbol{x}\| = \sqrt{1^2 + 1^2} = \sqrt{2}$. This is Pythagoras' theorem.

Properties. We have

1. If $a \in \mathbb{R}$, $\|a\boldsymbol{x}\| = |a|\|\boldsymbol{x}\|$.
2. Triangle inequality: if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ then $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.
3. Cauchy-Schwarz inequality: $|\boldsymbol{x} \cdot \boldsymbol{y}| \leqslant \|\boldsymbol{x}\|\|\boldsymbol{y}\|$.

If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we define their **distance** as $\boldsymbol{d(x, y)} = \|\boldsymbol{x} - \boldsymbol{y}\|$.

Exercise

Question 1

Let $\boldsymbol{x} = (1, 1)$, $\boldsymbol{y} = (-1, 2)$, $\boldsymbol{a} = (1, 2, 3)$, $\boldsymbol{b} = (-1, 0, 2)$. Calculate, if possible,

1. $\boldsymbol{x} + \boldsymbol{y}$
2. $2 \cdot \boldsymbol{y}$
3. $\boldsymbol{x} \cdot \boldsymbol{y}$
4. $\|\boldsymbol{x} - \boldsymbol{y}\|$
5. $\boldsymbol{x} + \boldsymbol{a}$
6. $\boldsymbol{a} + \boldsymbol{b}$
7. $\|\boldsymbol{a} - \boldsymbol{b}\|$
8. $|\boldsymbol{a} \cdot \boldsymbol{b}|$
9. $\boldsymbol{y} \cdot \boldsymbol{a}$

Matrices

A **matrix** $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is an $m \times n$ table of numbers, with m rows and n columns:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Examples. $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$, $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \in \mathbb{R}^{1\times 3}$, $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in \mathbb{R}^{3\times 1}$.

$\mathbf{0} \in \mathbb{R}^{m\times n}$ is the matrix whose entries are all zero.

If we have a dataset of n people with k numerical variables, we can represent the data as a matrix $A \in \mathbb{R}^{n\times k}$. Each row is a person, each column is a variable.

Operations

Given two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ we can define their sum or difference $\boldsymbol{A} \pm \boldsymbol{B}$ in the obvious way:

$$\boldsymbol{A} \pm \boldsymbol{B} = \begin{pmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1n} \pm b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2n} \pm b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{m2} & \cdots & a_{mn} \pm b_{mn} \end{pmatrix}$$

Examples.

- $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$
- $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \\ 3 \end{pmatrix}$

Properties. We have

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A},$$

$$\boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C},$$

and

$$\boldsymbol{A} - \boldsymbol{A} = \boldsymbol{0}.$$

Operations

We can do scalar multiplication: if $c \in \mathbb{R}$ and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ then

$$c\boldsymbol{A} = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{pmatrix}.$$

Examples.

- $2 \cdot \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$
- $-\frac{1}{2} \cdot \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ -\frac{3}{2} \end{pmatrix}$

Property. If $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}$ then

$$c(\boldsymbol{A} + \boldsymbol{B}) = c\boldsymbol{A} + c\boldsymbol{B}$$

.

Matrix × Vector Multiplication

If $\boldsymbol{x} \in \mathbb{R}^n$ we can see it as a matrix $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$: $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and a vector $\boldsymbol{x} \in \mathbb{R}^n$ we can multiply them:

$$\boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix} \in \mathbb{R}^{m\times 1}.$$

Notice that the ith entry of $\boldsymbol{A}\boldsymbol{x}$ is the inner product of the ith row of $\boldsymbol{A}$ with $\boldsymbol{x}$.

Matrix × Vector Multiplication

Examples.

- $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$
- $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ can't be done.
- $\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$

Exercise

Question 2

Calculate:

1. $\begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
2. $\begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} \left[2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]$

Linear Modeling is Matrix Multiplication

Suppose we have units $i = 1, \ldots, n$, a vector of outcomes $\boldsymbol{y} \in \mathbb{R}^n$, and k independent variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$. We want to approximate the outcome as a linear function of the independent variables. In other words, we want coefficients $\beta_1, \ldots, \beta_k \in \mathbb{R}$ such that

$$y_i \approx \beta_1 \boldsymbol{x}_{1i} + \cdots + \beta_k \boldsymbol{x}_{ki}.$$

We can create a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times k}$ whose columns are the k variables. If $\boldsymbol{\beta} \in \mathbb{R}^k$ is the vector of coefficients,

$$\boldsymbol{X}\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{x}_{11} & \cdots & \boldsymbol{x}_{k1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{x}_{1n} & \cdots & \boldsymbol{x}_{kn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \beta_1 \boldsymbol{x}_{11} + \cdots + \beta_k \boldsymbol{x}_{k1} \\ \vdots \\ \beta_1 \boldsymbol{x}_{1n} + \cdots + \beta_k \boldsymbol{x}_{kn} \end{pmatrix}.$$

Therefore, the objective is to minimize the distance between the outcomes $\boldsymbol{y}$ and the linear model $\boldsymbol{X}\boldsymbol{\beta}$, i.e.,

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|.$$

OLS means doing this. The answer is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$. We'll see what this means.

Matrix Multiplication

If $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n \times k}$ then we define $\boldsymbol{AB} \in \mathbb{R}^{m \times k}$ by

$$\begin{aligned}
\boldsymbol{AB} &= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & \cdots & a_{11}b_{1k} + \cdots + a_{1n}b_{nk} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & \cdots & a_{m1}b_{1k} + \cdots + a_{mn}b_{nk} \end{pmatrix}
\end{aligned}$$

We take each column $\boldsymbol{b}_{\bullet 1}, \ldots, \boldsymbol{b}_{\bullet k}$ of $\boldsymbol{B}$ and form $\begin{pmatrix} \boldsymbol{Ab}_{\bullet 1} & \cdots & \boldsymbol{Ab}_{\bullet k} \end{pmatrix}$.

Examples

- $\underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}_{2\times 2} \underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 1 & -2 & 0 \end{pmatrix}}_{2\times 3} = \underbrace{\begin{pmatrix} -1 & 2 & 0 \\ 1 & 2 & 3 \end{pmatrix}}_{2\times 3}$
- $\underbrace{\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}}_{1\times 3} \underbrace{\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}}_{3\times 1} = \underbrace{\begin{pmatrix} 2 \end{pmatrix}}_{1\times 1}$
- $\underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}}_{3\times 2} \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}}_{2\times 3} = \underbrace{\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{3\times 3}$
- $\underbrace{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}_{3\times 1} \underbrace{\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}}_{1\times 3} = \underbrace{\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ -1 & -2 & -3 \end{pmatrix}}_{3\times 3}$

Exercise

Question 3

Calculate:

1. $\begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}$
2. $\begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} \left[2 \begin{pmatrix} 1 & 0 & -1 \\ -1 & 2 & 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} \right]$

Properties

- **Associativity.** If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{m\times k}$ and $\boldsymbol{C} \in \mathbb{R}^{k\times l}$ then

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}).$$

- **Distributivity.** If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{m\times k}$ then

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}.$$

  If $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$, $\boldsymbol{C} \in \mathbb{R}^{m\times k}$ then

$$(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}.$$

- **Commutativity with scalars.** $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times k}$ and $c \in \mathbb{R}$ then

$$c\boldsymbol{AB} = \boldsymbol{A}(c\boldsymbol{B}).$$

Multiplication is Not Commutative

Example.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}, \qquad \boldsymbol{B} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{AB} = \begin{pmatrix} 2 & 3 \\ -2 & 2 \end{pmatrix}, \qquad \boldsymbol{BA} = \begin{pmatrix} 1 & 7 \\ -1 & 3 \end{pmatrix}$$

so $\boldsymbol{AB} \neq \boldsymbol{BA}$.

Transpose

The **transpose** of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the matrix $\boldsymbol{A}^\top \in \mathbb{R}^{n \times m}$ whose rows are the columns of $\boldsymbol{A}$.

Sometimes people write $\boldsymbol{A}'$ instead of $\boldsymbol{A}^\top$.

Examples.

- $\begin{pmatrix} 1 \\ 2 \end{pmatrix}^\top = \begin{pmatrix} 1 & 2 \end{pmatrix}$
- $\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \end{pmatrix}^\top = \begin{pmatrix} 0 & -1 \\ 1 & 0 \\ 2 & 1 \end{pmatrix}$
- $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^\top = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$

Properties of the Transpose

- $(\boldsymbol{A}^\top)^\top = \boldsymbol{A}$.
- $(\boldsymbol{A} + \boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$.
- $(c\boldsymbol{A})^\top = c\boldsymbol{A}^\top$.
- $(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$.

Notice that if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ then $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{y}$:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^\top \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + \cdots + x_n y_n.$$

Exercise

Question 4

Take $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \\ 1 & 1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 2 \end{pmatrix}$. Calculate:

1. $\boldsymbol{A}\boldsymbol{B}$
2. $\boldsymbol{B}^\top \boldsymbol{A}^\top$
3. $\boldsymbol{A} + 2\boldsymbol{B}^\top$
4. $\boldsymbol{A}^\top + \boldsymbol{B}$