Calculus Intuition for Machine Learning 2023

Lecture 1: Linear Algebra

Tim van Erven

- This lecture covers **crucial foundations** for all machine learning techniques! (revisit these slides during the next 5 weeks)

Poll

This lecture will be about **matrices** and **vectors**.

Q. Who has seen matrices and vectors before?

(Except for 3Blue1Brown videos.)

Outline

Intro: machine learning concepts for linear regression

Vectors: addition, scaling, length

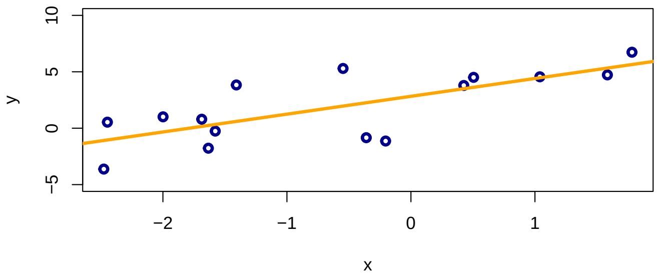
Matrices: addition, scaling, transpose

Vectors and matrices: multiplication

Identity matrix and matrix inverse

Reading the Least Squares Formula

Linear Regression with 1 Feature



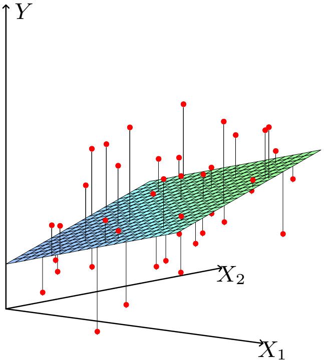
$n = 15$ data points: $\begin{pmatrix} y_1 \\ x_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ x_2 \end{pmatrix}, \ldots, \begin{pmatrix} y_n \\ x_n \end{pmatrix}$

Fit a line to the data: $f(x) = \theta_0 + x\theta_1$

- Can be used to predict y for a new unseen x
- Example: predict weight y from height x

Linear Regression with Multiple Features

Image from Hastie, Tibshirani, and Friedman (2009)



$$f(x_1, \ldots, x_d) = \theta_0 + x_1\theta_1 + \ldots + x_d\theta_d$$

- More information: d features instead of 1
- Example: predict weight y from $x = $ (height, age)

Linear Regression with Multiple Features

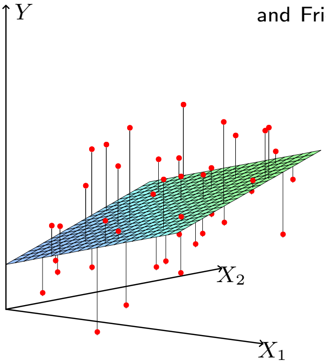


Image from Hastie, Tibshirani, and Friedman (2009)

$$f(x_1, \ldots, x_d) = \theta_0 + x_1\theta_1 + \ldots + x_d\theta_d$$

Each x_i with d features is a vector:

$$x_1 = \begin{pmatrix} x_{1,1} \\ \vdots \\ x_{1,d} \end{pmatrix}, \qquad x_2 = \begin{pmatrix} x_{2,1} \\ \vdots \\ x_{2,d} \end{pmatrix}, \qquad \ldots, \qquad x_n = \begin{pmatrix} x_{n,1} \\ \vdots \\ x_{n,d} \end{pmatrix}$$

How to represent this on a computer

Put all the responses together in one **vector** y:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

How to represent this on a computer

Put all the responses together in one **vector** y:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- How can we collect the feature vectors $x_1, \ldots, x_n$, which are already vectors?

[1] Programmers call this an 'array'

How to represent this on a computer

Put all the responses together in one **vector** y:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- How can we collect the feature vectors $x_1, \ldots, x_n$, which are already vectors?

Put all the feature vectors together in one $n \times d$ **matrix**[1] X:

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,d} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,d} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \ldots & x_{n,d} \end{pmatrix}$$

[1] Programmers call this an ‘array’

How to represent this on a computer

Put all the responses together in one **vector** y:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

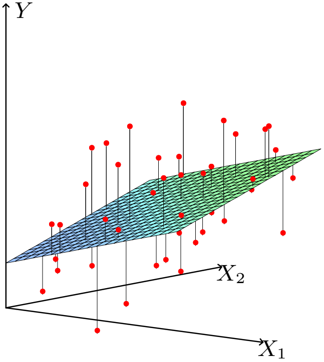
- How can we collect the feature vectors $x_1, \ldots, x_n$, which are already vectors?

Put all the feature vectors together in one $n \times d$ **matrix**[1] X:

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,d} \\ \boxed{x_{2,1} \quad x_{2,2} \quad \ldots \quad x_{2,d}} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \ldots & x_{n,d} \end{pmatrix} \quad x_2$$

[1] Programmers call this an 'array'

The Least Squares Method

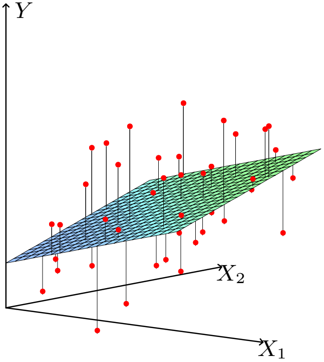


$$f(x_1, \ldots, x_d) = \theta_0 + x_1\theta_1 + \ldots + x_d\theta_d$$

Choose coefficients $\theta_0, \ldots, \theta_d$ to minimize the **sum of squared errors** on the data:

$$\sum_{i=1}^{n} \Big(y_i - f(x_i)\Big)^2$$

The Least Squares Method



$$f(x_1, \ldots, x_d) = \theta_0 + x_1\theta_1 + \ldots + x_d\theta_d$$

Choose coefficients $\theta_0, \ldots, \theta_d$ to minimize the **sum of squared errors** on the data:

$$\sum_{i=1}^{n} \Big(y_i - f(x_i)\Big)^2 = \sum_{i=1}^{n} \Big(y_i - (\theta_0 + x_{i,1}\theta_1 + \ldots + x_{i,d}\theta_d)\Big)^2$$

The Least Squares Formula

The data can be pre-processed such that $\theta_0 = 0$. Put the other coefficients together in a vector:

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix}$$

The Least Squares Formula

The data can be pre-processed such that $\theta_0 = 0$. Put the other coefficients together in a vector:

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix}$$

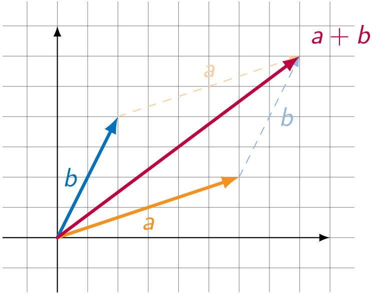
Then the **least squares solution** can be computed from the data by:

$$\theta = (X^\intercal X)^{-1} X^\intercal y$$

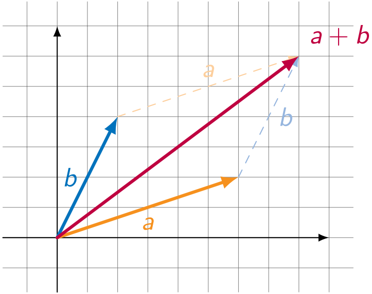
- Rest of the lecture: explain how to read this formula

Outline

Vectors: addition



Vectors: addition

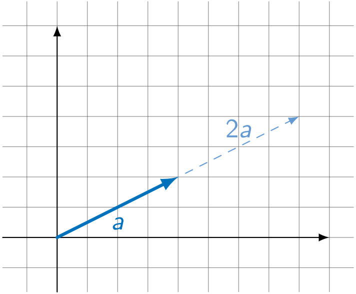


Addition: $a + b$

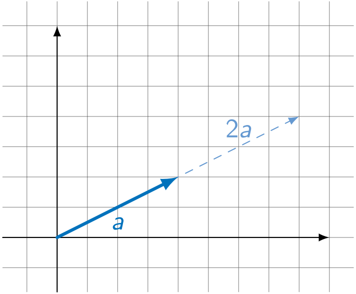
$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_d + a_d \end{pmatrix}$$

- Add coordinates separately

Vectors: scaling



Vectors: scaling

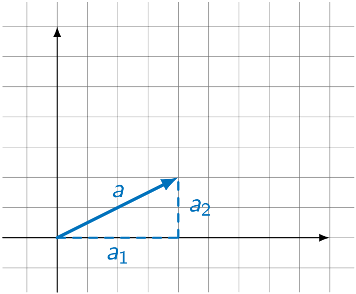


Scaling: $c \times a$ for some number c

$$c \times \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_d \end{pmatrix}$$

- Scale coordinates separately

Vectors: length



Length: $\|a\|$

$$\left\| \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} \right\| = \sqrt{a_1^2 + a_2^2 + \dots a_d^2}$$

- For $d = 2$ this is the Pythagorean theorem

Outline

Matrices: scaling

An $n \times d$ matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ a_{2,1} & a_{2,2} & \dots & a_{2,d} \\ & \vdots & & \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix}$$

- NB Vectors are $n \times 1$ matrices

Matrices: scaling

An $n \times d$ matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ a_{2,1} & a_{2,2} & \dots & a_{2,d} \\ & \vdots & & \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix}$$

Scaling: $c \times A$

$$c \times A = \begin{pmatrix} ca_{1,1} & ca_{1,2} & \dots & ca_{1,d} \\ ca_{2,1} & ca_{2,2} & \dots & ca_{2,d} \\ & \vdots & & \\ ca_{n,1} & ca_{n,2} & \dots & ca_{2,d} \end{pmatrix}$$

- Per coordinate, like for vectors

Matrices: addition

An $n \times d$ matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ a_{2,1} & a_{2,2} & \dots & a_{2,d} \\ & \vdots & & \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix}$$

Matrices: addition

An $n \times d$ matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ a_{2,1} & a_{2,2} & \dots & a_{2,d} \\ & \vdots & & \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix}$$

Addition: $A + B$

$$A + B = \begin{pmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} & \dots & a_{1,d} + b_{1,d} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} & \dots & a_{2,d} + b_{2,d} \\ & \vdots & & \\ a_{n,1} + b_{n,1} & a_{n,2} + b_{n,2} & \dots & a_{n,d} + b_{n,d} \end{pmatrix}$$

- Per coordinate, like for vectors

Matrices: transpose

An $n \times d$ matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ a_{2,1} & a_{2,2} & \dots & a_{2,d} \\ & \vdots & & \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix}$$

Transpose $A^\intercal$: Flip a matrix along its diagonal

A

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

Image by Lucas Vieira,

https://commons.wikimedia.org/w/index.php?curid=21897854

Outline

Vectors: Multiplication

The Inner Product $\langle a, b \rangle$ between two vectors:

$$\langle a, b \rangle = a_1 b_1 + a_2 b_2 + \ldots + a_d b_d$$

- E.g. $f(x) = \theta_0 + x_1\theta_1 + \ldots + x_d\theta_d = \theta_0 + \langle x, \theta \rangle$

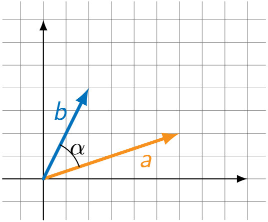
Vectors: Multiplication

The Inner Product $\langle a, b\rangle$ between two vectors:

$$\langle a, b\rangle = a_1 b_1 + a_2 b_2 + \ldots + a_d b_d$$

- E.g. $f(x) = \theta_0 + x_1\theta_1 + \ldots + x_d\theta_d = \theta_0 + \langle x, \theta\rangle$

Interpretation:



If α is the angle between the two vectors, then

$$\langle a, b\rangle = \cos(\alpha) \times \|a\| \times \|b\|$$

Matrices: Multiplication

$\boxed{B\text{: } d \textbf{ rows}, m \textbf{ columns}}$

$$\begin{pmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,m} \\ b_{2,1} & b_{2,2} & \dots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{d,1} & b_{d,2} & \dots & b_{d,m} \end{pmatrix}$$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ a_{2,1} & a_{2,2} & \dots & a_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix} \qquad \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ c_{2,1} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \dots & c_{n,m} \end{pmatrix}$$

$\boxed{A\text{: } n \textbf{ rows}, d \textbf{ columns}}$ $\qquad$ $\boxed{C = A \times B\text{: } n \textbf{ rows}, m \textbf{ columns}}$

Matrices: Multiplication

$$\begin{pmatrix} \boxed{\begin{matrix} b_{1,1} \\ b_{2,1} \\ \vdots \\ b_{d,1} \end{matrix}} & \begin{matrix} b_{1,2} \\ b_{2,2} \\ \vdots \\ b_{d,2} \end{matrix} & \begin{matrix} \dots \\ \dots \\ \ddots \\ \dots \end{matrix} & \begin{matrix} b_{1,m} \\ b_{2,m} \\ \vdots \\ b_{d,m} \end{matrix} \end{pmatrix}$$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ \hline a_{2,1} & a_{2,2} & \dots & a_{2,d} \\ \hline \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix} \qquad \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ \boxed{c_{2,1}} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \dots & c_{n,m} \end{pmatrix}$$

Let a_i^T be the i-th row of A and b_j the j-th column of B. Then

$$c_{i,j} = \langle a_i, b_j \rangle$$

Matrices: Multiplication

$$\begin{pmatrix} \boxed{\begin{matrix} b_{1,1} \\ b_{2,1} \\ \vdots \\ b_{d,1} \end{matrix}} & \begin{matrix} b_{1,2} \\ b_{2,2} \\ \vdots \\ b_{d,2} \end{matrix} & \begin{matrix} \dots \\ \dots \\ \ddots \\ \dots \end{matrix} & \begin{matrix} b_{1,m} \\ b_{2,m} \\ \vdots \\ b_{d,m} \end{matrix} \end{pmatrix}$$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,d} \\ \multicolumn{4}{c}{\boxed{a_{2,1} \quad a_{2,2} \quad \dots \quad a_{2,d}}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,d} \end{pmatrix} \qquad \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ \boxed{c_{2,1}} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \dots & c_{n,m} \end{pmatrix}$$

Let $a_i^\intercal$ be the i-th row of A and b_j the j-th column of B. Then

$$c_{i,j} = \langle a_i, b_j \rangle$$

Example: For any two vectors $v^\intercal w = \langle v, w \rangle$

Examples

Inner product:

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\rangle =$$

Examples

Inner product:

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\rangle = 1 \times 3 + 2 \times 4 = 12$$

Examples

Inner product:

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\rangle = 1 \times 3 + 2 \times 4 = 12$$

Matrix multiplication I:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 1 & 1 \\ 0 & 3 \end{pmatrix}$$

$$=$$

Examples

Inner product:

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\rangle = 1 \times 3 + 2 \times 4 = 12$$

Matrix multiplication I:

$$\begin{aligned} &\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 1 & 1 \\ 0 & 3 \end{pmatrix} \\ &\qquad = \begin{pmatrix} (1 \times -1 + 2 \times 1 + 3 \times 0) & (1 \times 2 + 2 \times 1 + 3 \times 3) \\ (2 \times -1 + 3 \times 1 + 4 \times 0) & (2 \times 2 + 3 \times 1 + 4 \times 3) \end{pmatrix} \\ &\qquad = \begin{pmatrix} 1 & 13 \\ 1 & 19 \end{pmatrix} \end{aligned}$$

Examples

Inner product:

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\rangle = 1 \times 3 + 2 \times 4 = 12$$

Matrix multiplication II:

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 1 & 1 \\ 0 & 3 \end{pmatrix} \quad =$$

Examples

Inner product:

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\rangle = 1 \times 3 + 2 \times 4 = 12$$

Matrix multiplication II:

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 1 & 1 \\ 0 & 3 \end{pmatrix} \quad = \quad \textbf{Not possible!}$$

(because dimensions do not match)

Outline

Intro: machine learning concepts for linear regression

Vectors: addition, scaling, length

Matrices: addition, scaling, transpose

Vectors and matrices: multiplication

Identity matrix and matrix inverse

Reading the Least Squares Formula

Identity Matrix

Is there a matrix which behaves like the number 1?
Yes, the **identity matrix** I!

B: d **rows**, m **columns**

$$\begin{pmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,m} \\ b_{2,1} & b_{2,2} & \dots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{d,1} & b_{d,2} & \dots & b_{d,m} \end{pmatrix}$$

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \qquad \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ c_{2,1} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \dots & c_{n,m} \end{pmatrix}$$

I: d **rows**, d **columns**

$C = I \times B$: d **rows**, m **columns**

Identity Matrix

Is there a matrix which behaves like the number 1?
Yes, the **identity matrix** I!

$$\begin{pmatrix} \boxed{\begin{matrix} b_{1,1} \\ b_{2,1} \\ \vdots \\ b_{d,1} \end{matrix}} & \begin{matrix} b_{1,2} \\ b_{2,2} \\ \vdots \\ b_{d,2} \end{matrix} & \begin{matrix} \dots \\ \dots \\ \ddots \\ \dots \end{matrix} & \begin{matrix} b_{1,m} \\ b_{2,m} \\ \vdots \\ b_{d,m} \end{matrix} \end{pmatrix}$$

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \boxed{0 \quad 1 \quad \dots \quad 0} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \qquad \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ \boxed{c_{2,1}} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \dots & c_{n,m} \end{pmatrix}$$

Identity Matrix

Is there a matrix which behaves like the number 1?
Yes, the **identity matrix** I!

$$\begin{pmatrix} \boxed{\begin{matrix} b_{1,1} \\ b_{2,1} \\ \vdots \\ b_{d,1} \end{matrix}} & \begin{matrix} b_{1,2} \\ b_{2,2} \\ \vdots \\ b_{d,2} \end{matrix} & \begin{matrix} \dots \\ \dots \\ \ddots \\ \dots \end{matrix} & \begin{matrix} b_{1,m} \\ b_{2,m} \\ \vdots \\ b_{d,m} \end{matrix} \end{pmatrix}$$

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \boxed{0 \quad 1 \quad \dots \quad 0} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \qquad \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ \boxed{c_{2,1}} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \dots & c_{n,m} \end{pmatrix}$$

$$c_{i,j} = b_{i,j} \quad \Longrightarrow \quad I \times B = B$$

Identity Matrix

Is there a matrix which behaves like the number 1?
Yes, the **identity matrix** I!

$$\begin{pmatrix} \boxed{\begin{matrix} b_{1,1} \\ b_{2,1} \\ \vdots \\ b_{d,1} \end{matrix}} & \begin{matrix} b_{1,2} \\ b_{2,2} \\ \vdots \\ b_{d,2} \end{matrix} & \begin{matrix} \dots \\ \dots \\ \ddots \\ \dots \end{matrix} & \begin{matrix} b_{1,m} \\ b_{2,m} \\ \vdots \\ b_{d,m} \end{matrix} \end{pmatrix}$$

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \boxed{0 \quad 1 \quad \dots \quad 0} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \qquad \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ \boxed{c_{2,1}} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & \dots & c_{n,m} \end{pmatrix}$$

$$c_{i,j} = b_{i,j} \qquad \Longrightarrow \qquad I \times B = B$$

Likewise: $B \times I = B$

Matrix Inverse

A square matrix A has an **inverse** A^{-1} if

$$A^{-1} \times A \quad = \quad I \quad = \quad A \times A^{-1}$$

- Think of A^{-1} as "dividing by A"
- The inverse is always unique (if it exists)

Matrix Inverse

A square matrix A has an **inverse** A^{-1} if

$$A^{-1} \times A \quad = \quad I \quad = \quad A \times A^{-1}$$

- Think of A^{-1} as "dividing by A"
- The inverse is always unique (if it exists)

Example:

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

Outline

Reading the Least Squares Formula

$$\theta = (X^\intercal X)^{-1} X^\intercal y$$

- $X^\intercal$ is the **transpose** of X
- $(X^\intercal X)^{-1}$ is the **inverse** of the matrix $X^\intercal X$
- All other operations are **matrix multiplications**, but we omit the $\times$ symbol. E.g. $X^\intercal X = X^\intercal \times X$

Preparation for Next Lecture

The lecture on Thursday will be about **derivatives and optimization**.

To prepare, watch the following video:

Math&Stuff - The Intuitive Concept of a Derivative

References I

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning*. 2nd ed. Springer.