Center for Statistics and the Social Sciences
Math Camp 2022

Lecture 2, Part 1: Matrix Algebra

Jess Kunke & Erin Lipman

Department of Statistics
University of Washington

Tue 13 Sep 2022

Motivation

Matrices are arrays of numbers

- Matrices make solving systems of linear equations more efficient
- Matrix algebra provides concise notation and rules for manipulating matrices
- Applications:
  - Linear regression/least squares
  - Working with data sets (rows and columns)

Outline for today

Definitions and terminology

Matrix operations: addition, multiplication, division

Solving systems of linear equations with matrices

Application: Linear regression and least squares

Outline for today

Definitions and terminology

Matrix operations: addition, multiplication, division

Solving systems of linear equations with matrices

Application: Linear regression and least squares

Definitions & Notation

What is a matrix?

A **matrix** is an array of numbers in a rectangular form.

Examples:

$$A = \begin{bmatrix} 1 & 2 & 6 & 4 \\ 5 & 8 & 12 & 8 \\ 4 & 3 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 3 & 2 \\ 1 & 2 & 4 \end{bmatrix}.$$

Notes:

- These numbers are called the **elements** of the matrix
- Matrix dimensions ($n \times m$) are always listed as rows $\times$ columns (A is a 3×4 matrix and B is a 2×3 matrix)
- A is also sometimes written $A_{n \times m}$

Definitions & Notation

Matrix indexing

Matrix elements are indexed as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix},$$

where x_{ij} is the value in the ith row and the jth column of matrix X.

- Tip: make an L

Definitions & Notation

Special Matrices

A **vector** is a single row or column, i.e. a matrix in which one dimension equals 1.
Examples: What are the dimensions of these vectors?

$$F = \begin{bmatrix} 1 & 2 & 6 & 4 \end{bmatrix}, \quad G = \begin{bmatrix} 4 \\ 5 \\ 1 \end{bmatrix}$$

Definitions & Notation

Special Matrices

A **vector** is a single row or column, i.e. a matrix in which one dimension equals 1.
Examples: What are the dimensions of these vectors?

$$F = \begin{bmatrix} 1 & 2 & 6 & 4 \end{bmatrix}, \quad G = \begin{bmatrix} 4 \\ 5 \\ 1 \end{bmatrix}$$

A **square** matrix has the same number of rows and columns.
Example: This is a 2x2 square matrix.

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

Definitions & Notation

Special Matrices

A **symmetric** matrix has elements such that $x_{ij} = x_{ji}$.
Example:

$$\begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 3 \\ 5 & 3 & 7 \end{bmatrix}$$

Definitions & Notation

Special Matrices

A **symmetric** matrix has elements such that $x_{ij} = x_{ji}$.
Example:

$$\begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 3 \\ 5 & 3 & 7 \end{bmatrix}$$

A symmetric matrix must also be a square matrix.

Definitions & Notation

Special Matrices

A **diagonal** matrix is a matrix that is zero everywhere except on the diagonal. Where the diagonal is defined as all elements for which the row number is equal to the column number $\{(1,1),(2,2),(3,3),...\}$. For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

is a diagonal matrix.

Definitions & Notation

Special Matrices

A **diagonal** matrix is a matrix that is zero everywhere except on the diagonal. Where the diagonal is defined as all elements for which the row number is equal to the column number $\{(1,1),(2,2),(3,3),...\}$. For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

is a diagonal matrix.

A special case of a diagonal matrix is the **identity** matrix. Its diagonal elements are all ones.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

These are very useful! We'll see a little more about them later.

Outline for today

Definitions and terminology

Matrix operations: addition, multiplication, division

Solving systems of linear equations with matrices

Application: Linear regression and least squares

Matrix Operations
Basic Operations

Matrix Equality: Two matrices A, B are equal if and only if, for all elements, each $a_{ij} = b_{ij}$. (Note: this means they must have the same dimensions.)

Matrix Operations
Basic Operations

Matrix Equality: Two matrices A, B are equal if and only if, for all elements, each $a_{ij} = b_{ij}$. (Note: this means they must have the same dimensions.)

Matrix Transpose: The **transpose** of a matrix is found by interchanging the corresponding rows and columns of a matrix. The first row becomes the first column, the second row becomes the second column, etc. The dimensions are then switched and the element a_{ij} becomes the element a_{ji}. The transposed matrix is often denoted A^t (or A'). You can find the transpose of a matrix in `R` by using the `t()` function.

Matrix Operations

Basic Operations

Matrix Equality: Two matrices A, B are equal if and only if, for all elements, each $a_{ij} = b_{ij}$. (Note: this means they must have the same dimensions.)

Matrix Transpose: The **transpose** of a matrix is found by interchanging the corresponding rows and columns of a matrix. The first row becomes the first column, the second row becomes the second column, etc. The dimensions are then switched and the element a_{ij} becomes the element a_{ji}. The transposed matrix is often denoted A^t (or A'). You can find the transpose of a matrix in `R` by using the `t()` function.

$$\text{Example:} \quad A = \begin{bmatrix} 1 & 2 & 6 \\ 3 & 5 & 9 \end{bmatrix} \quad A^t = ?$$

Matrix Operations
Addition & Subtraction

Two matrices can be added or subtracted only if their dimensions are the same (both rows and columns). The corresponding elements are then added or subtracted.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix}$$

Matrix Operations

Addition & Subtraction

Two matrices can be added or subtracted only if their dimensions are the same (both rows and columns). The corresponding elements are then added or subtracted.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix}$$

Example:

$$\begin{bmatrix} 1 & 2 & 6 \\ 3 & 5 & 9 \end{bmatrix} - \begin{bmatrix} 1 & 3 & 8 \\ 6 & 9 & 6 \end{bmatrix} =?$$

Matrix Operations

Scalar Multiplication

To multiply a matrix by a **scalar** (a real number; any $a \in \mathbb{R}$), multiply each element by that number.

Example: Find $3A$ if

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 6 & 9 & 6 \end{bmatrix}.$$

Matrix Operations

Multiplication Examples

What about multiplying a matrix by another matrix?

Let's think about regular numbers first (also called **scalars**):

- We can multiply numbers, e.g. $4 \cdot 5 = 20$
- Multiplication is commutative, meaning order doesn't matter:

$$4 \cdot 5 = 5 \cdot 4 = 20$$

Matrix Operations

Multiplication Examples

What about multiplying a matrix by another matrix?

Let's think about regular numbers first (also called **scalars**):

- We can multiply numbers, e.g. $4 \cdot 5 = 20$
- Multiplication is commutative, meaning order doesn't matter:

$$4 \cdot 5 = 5 \cdot 4 = 20$$

Unlike scalars/numbers, you can't always multiply two matrices, and multiplication is not in general commutative!

Matrix Operations

Multiplying two vectors

You can multiply a **row vector** times a **column vector** as follows:

Matrix Operations

Multiplying two vectors

You can multiply a **matrix** times a **column vector** as follows:

What do you notice has to be true about the dimensions in order to multiply them?

Matrix Operations

Multiplication Examples

Two matrices $A_{n_A \times m_A}$ and $B_{n_B \times m_B}$ can be multiplied only if the number of columns of the first matrix, m_A, equals the number of rows of the second matrix, n_B, i.e. the "inside numbers".

The resulting matrix $A \cdot B$ (also written AB) has n_A rows and m_B columns, i.e. the "outside numbers".

Matrix Operations

Multiplication Examples

Two matrices $A_{n_A \times m_A}$ and $B_{n_B \times m_B}$ can be multiplied only if the number of columns of the first matrix, m_A, equals the number of rows of the second matrix, n_B, i.e. the "inside numbers".

The resulting matrix $A \cdot B$ (also written AB) has n_A rows and m_B columns, i.e. the "outside numbers".

Example:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

- What are the dimensions of A and B?
- Can you compute AB? What dimension will it have?
- Can you compute BA? What dimension will it have?

Matrix Operations

Multiplication Examples

To compute AB, we find each element $(ab)_{ij}$ by summing the cross products of the ith row of A and the jth column of B:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$AB = \begin{bmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} + a_{13} \cdot b_{31} & a_{11} \cdot b_{12} + a_{12} \cdot b_{22} + a_{13} \cdot b_{32} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} + a_{23} \cdot b_{31} & a_{21} \cdot b_{12} + a_{22} \cdot b_{22} + a_{23} \cdot b_{32} \end{bmatrix}$$

Matrix Operations

Multiplication Examples

To compute AB, we find each element $(ab)_{ij}$ by summing the cross products of the ith row of A and the jth column of B:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$AB = \begin{bmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} + a_{13} \cdot b_{31} & a_{11} \cdot b_{12} + a_{12} \cdot b_{22} + a_{13} \cdot b_{32} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} + a_{23} \cdot b_{31} & a_{21} \cdot b_{12} + a_{22} \cdot b_{22} + a_{23} \cdot b_{32} \end{bmatrix}$$

Example:

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 6 & 9 & 6 \\ 2 & 1 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 9 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}$$

$$AB = \begin{bmatrix} 1 \cdot 3 + 3 \cdot 2 + 8 \cdot 3 & 1 \cdot 9 + 3 \cdot 1 + 8 \cdot 2 \\ 6 \cdot 3 + 9 \cdot 2 + 6 \cdot 3 & 6 \cdot 9 + 9 \cdot 1 + 6 \cdot 2 \\ 2 \cdot 3 + 1 \cdot 2 + 3 \cdot 3 & 2 \cdot 9 + 1 \cdot 1 + 3 \cdot 2 \end{bmatrix} = \begin{bmatrix} 33 & 28 \\ 54 & 75 \\ 17 & 25 \end{bmatrix}$$

Matrix Multiplication

Order Matters

Can we compute BA in the previous example? Why or why not? If we can, what is the dimension of BA?

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 6 & 9 & 6 \\ 2 & 1 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 9 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}$$

Matrix Operations

Matrix multiplication

What about division? Again, let's first think about scalars.

The multiplicative **identity** for the real numbers is the number d that if you multiply any real number a by d, you get a. There is one identity for all the real numbers. What is d?

Matrix Operations

Matrix multiplication

What about division? Again, let's first think about scalars.

The multiplicative **identity** for the real numbers is the number d that if you multiply any real number a by d, you get a. There is one identity for all the real numbers. What is d?

For each real number a, its multiplicative **inverse** if it exists is the number a^{-1} such that the product of a and a^{-1} is the identity, i.e. $a \cdot a^{-1} = d$. What is a formula for a^{-1} in terms of a?

Matrix Operations

Inverse

This way of framing multiplication might feel unnecessarily abstract if you only need to deal with real numbers, but it will help us generalize the idea of multiplication for matrices.

The **inverse** of a matrix $A_{n\times n}$ is the matrix $A^{-1}_{n\times n}$ that satisfies

$$A \cdot A^{-1} = I,$$

where is the **identity matrix**. For example, $I_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Like the multiplicative identity, any matrix multiplied by I is itself:

$$A \times I = I \times A = A.$$

Matrix Operations

Determinant

How do we find the inverse? How do we know if the inverse exists?

The **determinant** of a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is $D(A) = a \cdot d - b \cdot c$.

Matrix Operations

Determinant

How do we find the inverse? How do we know if the inverse exists?

The **determinant** of a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is $D(A) = a \cdot d - b \cdot c$.

- If $D(A) = 0$, A^{-1} does not exist. A is **singular**.

Matrix Operations

Determinant

How do we find the inverse? How do we know if the inverse exists?

The **determinant** of a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is $D(A) = a \cdot d - b \cdot c$.

- If $D(A) = 0$, A^{-1} does not exist. A is **singular**.
- If $D(A) \neq 0$, A^{-1} exists. A is **nonsingular**.

Matrix Operations

Determinant

Examples: for each matrix, compute the determinant and state whether the inverse exists.

$$C = \begin{bmatrix} 4 & 12 \\ 3 & 6 \end{bmatrix}$$

$$M = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$$

Matrix Operations

Determinant: geometric interpretation

We can also use geometric reasoning to determine whether the inverse exists!

$$C = \begin{bmatrix} 4 & 12 \\ 3 & 6 \end{bmatrix}, \qquad M = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$$

Matrix Operations

Inverse Example

If the inverse A^{-1} exists for $A_{2\times 2}$, we can compute it.

For higher dimensions, let a computer do it.

- The function `solve()` computes matrix inverses in `R`.
- Inverting big matrices can take **a lot** of computing power.

Let's see how to compute inverses for a 2x2 matrix!

Matrix Operations

Computing the inverse

The inverse of a matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is given by

$$A^{-1} = \frac{1}{D(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \quad \text{where } D(A) = ad - bc.$$

Example: What is the inverse of $A = \begin{bmatrix} 4 & 12 \\ 3 & 6 \end{bmatrix}$?

Outline for today

Definitions and terminology

Matrix operations: addition, multiplication, division

Solving systems of linear equations with matrices

Application: Linear regression and least squares

Linear Equations

Let's go back to thinking about systems of two equations:

$$ax + by = g$$

$$cx + dy = f$$

Linear Equations

Let's go back to thinking about systems of two equations:

$$ax + by = g$$

$$cx + dy = f$$

Previously we solved this system by eliminating the y variable, solving for x, and then substituting back in for y.

Linear Equations

Let's go back to thinking about systems of two equations:

$$ax + by = g$$

$$cx + dy = f$$

Previously we solved this system by eliminating the y variable, solving for x, and then substituting back in for y.

No we can write this system in matrix notation:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad z = \begin{bmatrix} x \\ y \end{bmatrix}, \quad w = \begin{bmatrix} g \\ f \end{bmatrix}$$

Solving our system of equations is the same as solving for z in this matrix equation, where A and z are multiplied using matrix multiplication:

$$Az = w$$

Linear Equations

Examples

Solving our system of equations is the same as solving for z in the matrix equation:

So how do we solve for z?

$$\begin{aligned} A \cdot z &= w \\ A^{-1} \cdot A \cdot z &= A^{-1} \cdot w && [\text{Left-multiply by } A^{-1}] \\ I \cdot z &= A^{-1} \cdot w && [A^{-1} \times A = I] \\ z &= A^{-1} \cdot w. \end{aligned}$$

The solution to our system is $z = A^{-1} \cdot w = \begin{bmatrix} x \\ y \end{bmatrix}$.

Linear Equations

Example

Let's try a concrete example together:

$$2x + y = 1$$

$$4x + 3y = 8$$

Outline for today

Definitions and terminology

Matrix operations: addition, multiplication, division

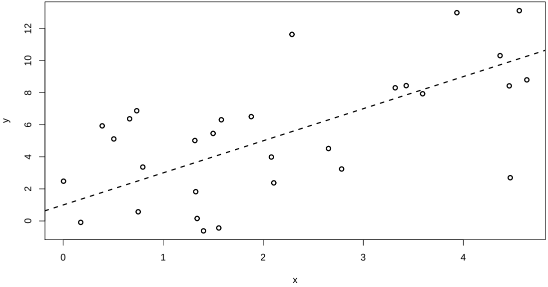
Solving systems of linear equations with matrices

Application: Linear regression and least squares

Linear Regression and Least Squares

The goal of **linear regression** is to estimate the intercept and slope in a linear relationship between an independent variable or covariate X and a dependent variable or outcome, Y.

In other words, we want to fit a line through pairs of points (x_i, y_i) for observations $i = 1, \ldots, n$.



Least Squares

So how do we choose the dashed line? First let's suppose we knew the true x's and y's and create a mathematical model of the true relationship.

- In this setting we have an equation for each observation, so if we've got $n > 2$ observations we have n many equations with only two coefficients we want to estimate

In 2D, for each observation $i = 1, \ldots, n$ we can write the equation:

$$y_i = \beta_0 + \beta_1 x_i,$$

where

- x_i is the **true** (in practice unknown) value of the independent variable for obs i
- y_i is the **true** (in practice unknown) value of the response for obs i
- β_0 is the intercept
- β_1 is the slope

Least Squares

- The true values fall on this line and we don't know them
- If we knew the true values, we'd only need $n = 2$ observations, two pairs of x and y
- But we don't, so we want more observations to get a better estimate (we'll clarify this more on Thursday)

Least Squares

In 2D (one independent variable x)

Our model of what we think the true relationship is:

$$y_i = \beta_0 + \beta_1 x_i \text{ for each pair } i$$

In matrix notation:

$$y = X\beta = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & \dots \\ 1 & x_{n1} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- $y_{n\times 1}$ is the **response**.
- $X_{n\times 2}$ is the **design matrix**.

  Notice the column of 1's so that each observation's model includes a β_0.
- $\beta_{2\times 1}$ are the unknown **coefficients** we want to estimate.

Least Squares

More than one independent variable

Note: we can generalize to p many independent variables using

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{pi}$$

In matrix notation:

$$y = X\beta = \begin{bmatrix} y_1 \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & ... & x_{1p} \\ 1 & ... & ... & ... \\ 1 & x_{n1} & ... & x_{np} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ ... \\ \beta_p \end{bmatrix}$$

- $y_{n\times 1}$ is the **response**.
- $X_{n\times(p+1)}$ is the **design matrix**.
  Notice the column of 1's so that each observation's model includes a β_0.
- $\beta_{(p+1)\times 1}$ are the unknown **coefficients** we want to estimate.

Least Squares

In 2D (one independent variable x)

For now let's stick to 2D:

$$y_i = \beta_0 + \beta_1 x_i$$

In matrix notation:

$$y = X\beta = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & \dots \\ 1 & x_{n1} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- $y_{n\times 1}$ is the **response**.
- $X_{n\times 2}$ is the **design matrix**.

  Notice the column of 1's so that each observation's model includes a β_0.
- $\beta_{2\times 1}$ are the unknown **coefficients** we want to estimate.

Least Squares

Under the true model, if we knew the true x's and y's, we could find the true β_0 and β_1 as

$$\beta = \left(X^t \cdot X\right)^{-1} X^t y.$$

To see this:

Least Squares

But we don't know the true x's and y's! Instead think we have noisy estimates of them.

So if we plug in the **observed** x's and y's instead of the true ones, we no longer have exact equalities but approximations:

$$y_i \approx \beta_0 + \beta_1 x_i$$

In matrix notation:

$$y \approx X\beta$$

$$\begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} \approx \begin{bmatrix} 1 & x_{11} \\ 1 & \dots \\ 1 & x_{n1} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Least Squares

But we don't know the true x's and y's! Instead think we have noisy estimates of them.

So if we plug in the **observed** x's and y's instead of the true ones, we no longer have exact equalities but approximations:

$$y_i \approx \beta_0 + \beta_1 x_i$$

In matrix notation:

$$y \approx X\beta$$

$$\begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} \approx \begin{bmatrix} 1 & x_{11} \\ 1 & \dots \\ 1 & x_{n1} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Our approximation can be better or worse depending on how we choose β. How do we choose the "best" β?

Least Squares

One approach is **least squares**.

Least squares finds the line that minimizes the squared distance between the points and the line, i.e. makes

$$[y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi})]^2$$

as small as possible for all $i = 1, \ldots, n$.

Least Squares

One approach is **least squares**.

Least squares finds the line that minimizes the squared distance between the points and the line, i.e. makes

$$[y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi})]^2$$

as small as possible for all $i = 1, \ldots, n$.

(This is optimization! We'll see this tomorrow with derivatives)

Least squares estimates

The vector $\widehat{\beta}$ that minimizes the sum of the squared distances turns out to be

$$\widehat{\beta} = \left(X^t \cdot X\right)^{-1} X^t y.$$

- To get this we make the approximations equalities again by adding in **noise** (Thursday) and taking derivatives (Wednesday)

Least squares estimates

The vector $\widehat{\beta}$ that minimizes the sum of the squared distances turns out to be

$$\widehat{\beta} = \left(X^t \cdot X\right)^{-1} X^t y.$$

- To get this we make the approximations equalities again by adding in **noise** (Thursday) and taking derivatives (Wednesday)
- In statistics, once we have estimated a parameter we put a "hat" on it, e.g. $\widehat{\beta_0}$ is the estimate of the true parameter β_0.

Connection to lab today

Today in lab we'll try out a simple linear regression example in R