Lectures 8-10

Linear algebra: Dot products and Projections

DSC 40A, Fall 2025

Question 🤔

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!

If the direct link doesn't work, click the "🤔 Lecture Questions" link in the top right corner of dsc40a.com.

Agenda

- Recap: Simple linear regression and correlation.
- Connections to related models. ← HW3
- Dot products.
- Spans and projections.
- Normal equations

* HW2 due tonight
* HW3 released tonight
* HW1 grades will be released tonight (hopefully :))

Friday OH moved to HDSI 336!

Orthogonal projection

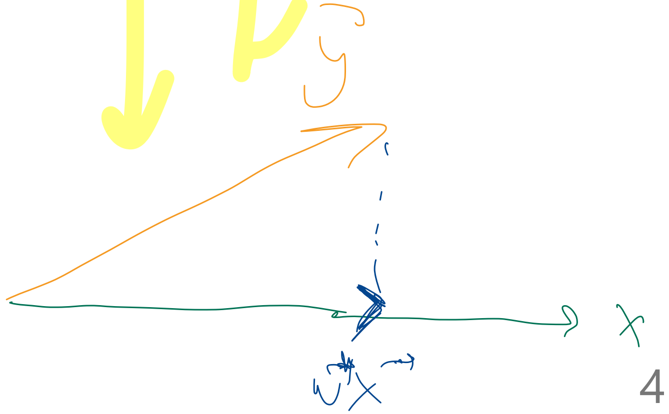
- **Question**: What vector in $\text{span}(\vec{x})$ is closest to $\vec{y}$?
- **Answer**: It is the vector $w^* \vec{x}$, where:

$$w^* = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}} = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}$$

- Note that w^* is the solution to a minimization problem, specifically, this one:

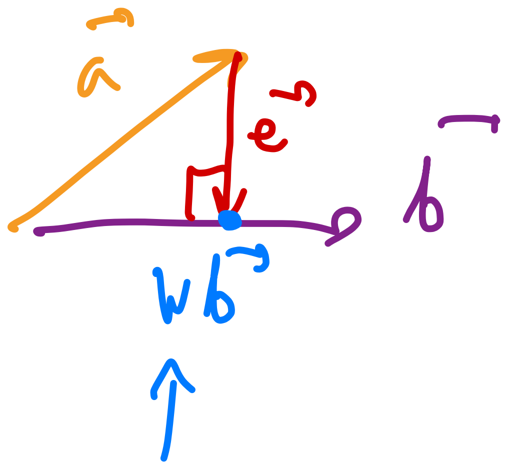
$$\text{error}(w) = \|\vec{e}\| = \|\vec{y} - w\vec{x}\|$$

- We call $w^* \vec{x}$ the **orthogonal projection of $\vec{y}$ onto** $\text{span}(\vec{x})$.
  - Think of $w^* \vec{x}$ as the "shadow" of $\vec{y}$.



Exercise

Let $\vec{a} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 9 \end{bmatrix}$.



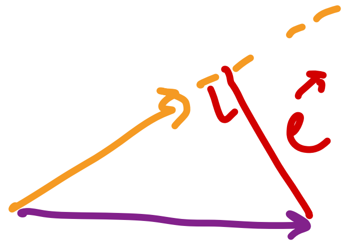
What is the orthogonal projection of $\vec{a}$ onto $\text{span}(\vec{b})$?

Your answer should be of the form $w^*\vec{b}$, where w^* is a scalar.

projecting $\vec{a}$ onto span$\{\vec{b}\}$

$$w^* = \frac{\vec{b}\cdot\vec{a}}{\vec{b}\cdot\vec{b}} = \frac{-1\cdot 5 + 9\cdot 2}{(-1)^2 + 9^2} = \frac{-5+18}{1+81} = \frac{13}{82}$$

projecting $\vec{b}$ onto span$\{\vec{a}\}$

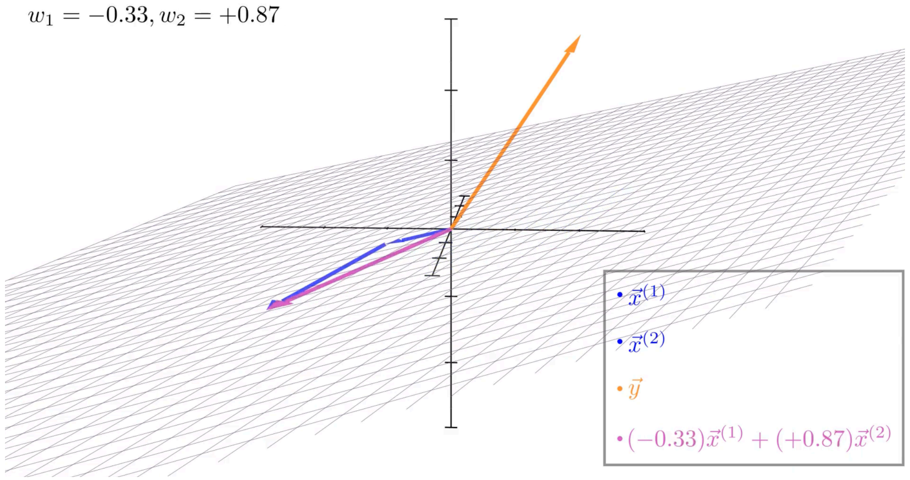


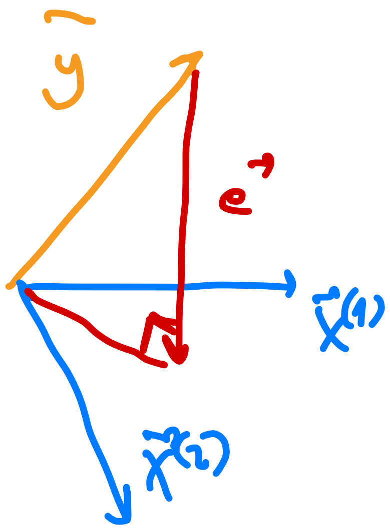
$$w^* = \frac{\vec{a}\cdot\vec{b}}{\vec{a}\cdot\vec{a}} = \frac{5\cdot(-1) + 2\cdot 9}{5^2+2^2} = \frac{13}{29}$$

feature index superscript

Moving to multiple dimensions

- Let's now consider three vectors, $\vec{y}$, $\vec{x}^{(1)}$, and $\vec{x}^{(2)}$, all in $\mathbb{R}^n$.
- **Question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
  - Vectors in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ are of the form $w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}$, where w_1, $w_2 \in \mathbb{R}$ are scalars.
- Before trying to answer, let's watch 🎥 **this animation that Jack, one of our previous tutors, made**.





Minimizing projection error in multiple dimensions

- **Question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
  - That is, what vector minimizes $\|\vec{e}\|$, where:

$$\vec{e} = \vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}$$

- **Answer**: It's the vector such that $w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}$ is **orthogonal** to $\vec{e}$.
- **Issue**: Solving for w_1 and w_2 in the following equation is difficult:

$$\left(w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}\right) \cdot \underbrace{\left(\vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}\right)}_{\vec{e}} = 0$$

Vector in the span$\{\vec{x}^{(1)}, \vec{x}^{(2)}\}$ is a linear comb. of $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$

Minimizing projection error in multiple dimensions

- It's hard for us to solve for w_1 and w_2 in:

$$\left(w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}\right) \cdot \underbrace{\left(\vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}\right)}_{\vec{e}} = 0$$

- **Observation**: All we really need is for $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ to individually be orthogonal to $\vec{e}$.
  - That is, it's sufficient for $\vec{e}$ to be orthogonal to the spanning vectors themselves.
- If $\vec{x}^{(1)} \cdot \vec{e} = 0$ and $\vec{x}^{(2)} \cdot \vec{e} = 0$, then:

$$(w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}) \cdot \vec{e} = (w_1 \vec{x}^{(1)}) \cdot \vec{e} + (w_2 \vec{x}^{(2)}) \vec{e}$$

$$= w_1 \underbrace{(\vec{x}_1 \cdot \vec{e})}_{=0} + w_2 \underbrace{(\vec{x}_2 \cdot \vec{e})}_{=0} = 0$$

Minimizing projection error in multiple dimensions

- **Question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
- **Answer**: It's the vector such that $\underbrace{w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}}_{h(x^{(1)}, x^{(2)})}$ is **orthogonal** to $\vec{e} = \vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}$.
- **Equivalently, it's the vector such that $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ are both orthogonal to $\vec{e}$:**

$$\boxed{\begin{aligned} \vec{x}^{(1)} \cdot \left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}\right) &= 0 \\ \vec{x}^{(2)} \cdot \underbrace{\left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}\right)}_{\vec{e}} &= 0 \end{aligned}}$$

solving these 2 equations will yield w_1^*, w_2^*

- This is a system of two equations, two unknowns (w_1 and w_2), but it still looks difficult to solve.

Now what?

- We're looking for the scalars w_1 and w_2 that satisfy the following equations:

$$\begin{aligned} \vec{x}^{(1)} \cdot \left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}\right) &= 0 \\ \vec{x}^{(2)} \cdot \underbrace{\left(\vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)}\right)}_{\vec{e}} &= 0 \end{aligned}$$

- In this example, we just have two spanning vectors, $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$.
- If we had any more, this system of equations would get extremely messy, extremely quickly.
- **Idea**: Rewrite the above system of equations **as a single equation, involving matrix-vector products**.

Matrices

Matrices

- An $n \times d$ **matrix** is a table of numbers with n rows and d columns.
- We use upper-case letters to denote matrices.

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix}$$

- Since A has two rows and three columns, we say $A \in \mathbb{R}^{2 \times 3}$.
- **Key idea**: Think of a matrix as **several column vectors, stacked next to each other**.

$$A = \left[\begin{bmatrix} 2 \\ -1 \end{bmatrix} \begin{bmatrix} 5 \\ 5 \end{bmatrix} \begin{bmatrix} 8 \\ -3 \end{bmatrix} \right]$$

Matrix addition and scalar multiplication

- We can add two matrices only if they have the same dimensions.

- Addition occurs elementwise:

$$C[i,j] = A[i,j] + B[i,j]$$

$$\underbrace{\begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix}}_{A} + \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} 3 & 7 & 11 \\ -1 & 6 & -1 \end{bmatrix}}_{C}$$

- Scalar multiplication occurs elementwise, too:

$$D[i,j] = cA[i,j]$$

$$\underbrace{2}_{c \,\in\, \mathbb{R}} \underbrace{\begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 4 & 10 & 16 \\ -2 & 10 & -6 \end{bmatrix}}_{D}$$

Matrix-matrix multiplication

- **Key idea**: We can multiply matrices A and B **if and only if**: AB

$$\boxed{\# \text{ columns in } A = \# \text{ rows in } B}$$

- If A is $n \times d$ and B is $d \times p$, then AB is $n \times p$.
- Example: If A is as defined below, what is $A^T A$?

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

$$A^T = \begin{bmatrix} 2 & -1 \\ 5 & 5 \\ 8 & -3 \end{bmatrix} \in \mathbb{R}^{3\times 2}$$

$$\underset{3\times 2 \quad 2\times 3}{A^T A} = \begin{bmatrix} 2\cdot 2 + (-1)\cdot(-1) & \cdot & \cdot \\ 5 & \cdot & \cdot \\ 15 & \cdot & \cdot \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

Question 🤔

Answer at q.dsc40a.com

Assume A, B, and C are all matrices. Select the **incorrect** statement below.

- A. $A(B + C) = AB + AC$.
- B. $A(BC) = (AB)C$.
- $\boxed{\text{C. } AB = BA.}$
- D. $(A + B)^T = A^T + B^T$.
- E. $(AB)^T = B^T A^T$.

example

$$\underset{5\times 7}{A} \cdot \underset{7\times 3}{B} = \underset{5\times 3}{C}$$

$$\underset{7\times 3}{B} \cdot \underset{5\times 7}{A} \rightarrow \text{product undefined!}$$

different

Matrix-vector multiplication

- A vector $\vec{v} \in \mathbb{R}^n$ is a matrix with n rows and 1 column.

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \in \mathbb{R}^{n \times 1}$$

- Suppose $A \in \mathbb{R}^{n \times d}$.
  - What must the dimensions of $\vec{v}$ be in order for the product $A\vec{v}$ to be valid?

    $\underset{n \times d}{A} \quad \underset{d \times 1}{v}$

    $\vec{v} \in \mathbb{R}^d$, column vec with d elements

  - What must the dimensions of $\vec{v}$ be in order for the product $\vec{v}^T A$ to be valid?

    $\underset{1 \times n}{\vec{v}^T} \cdot \underset{n \times d}{A}$

    $v \in \mathbb{R}^n$, v^T row vec with n elements

One view of matrix-vector multiplication

row perspective

- One way of thinking about the product $A\vec{v}$ is that it is **the dot product of $\vec{v}$ with every row of A.**
- Example: What is $A\vec{v}$?

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} 2 \\ -1 \\ -5 \end{bmatrix}$$

$$A\vec{v} \begin{bmatrix} [\vec{A}]_1 \cdot \vec{v} \\ [\vec{A}]_2 \cdot \vec{v} \end{bmatrix} = \begin{bmatrix} 2 \cdot 2 + 5 \cdot (-1) + 8 \cdot (-5) \\ -1 \cdot 2 + 5 \cdot (-1) + (-3) \cdot (-5) \end{bmatrix} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}$$

$\in \mathbb{R}^2$

Another view of matrix-vector multiplication

- Another way of thinking about the product $A\vec{v}$ is that it is **a linear combination of the columns of A, using the weights in $\vec{v}$.**
- Example: What is $A\vec{v}$?

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} 2 \\ -1 \\ -5 \end{bmatrix}$$

$$A\vec{v} = \underbrace{2\begin{bmatrix} 2 \\ -1 \end{bmatrix} - 1\begin{bmatrix} 5 \\ 5 \end{bmatrix} - 5\begin{bmatrix} 8 \\ -3 \end{bmatrix}}_{\text{linear combination of the cols of } A} = \begin{bmatrix} 4 \\ -2 \end{bmatrix} + \begin{bmatrix} -5 \\ -5 \end{bmatrix} + \begin{bmatrix} -40 \\ 15 \end{bmatrix} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}$$

Matrix-vector products create linear combinations of columns!

- **Key idea**: It'll be very useful to think of the matrix-vector product $A\vec{v}$ as a linear combination of the columns of A, using the weights in $\vec{v}$.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}$$

$$\downarrow$$

$$\boxed{A\vec{v} = v_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} + \ldots + v_d \begin{bmatrix} a_{1d} \\ a_{2d} \\ \vdots \\ a_{nd} \end{bmatrix}}$$

Spans and projections, revisited

Moving to multiple dimensions

- Let's now consider three vectors, $\vec{y}$, $\vec{x}^{(1)}$, and $\vec{x}^{(2)}$, all in $\mathbb{R}^n$.
- **Question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
  - That is, what values of w_1 and w_2 minimize $\|\vec{e}\| = \|\vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}\|$?

Find w_1^*, w_2^* such that

$$\begin{cases} \vec{x}^{(1)} \cdot \vec{e} = 0 \\ \vec{x}^{(2)} \cdot \vec{e} = 0 \end{cases}$$

Matrix-vector products create linear combinations of columns!

$$\vec{x}^{(1)} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} \qquad \vec{x}^{(2)} = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

$w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}$

$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

- Combining $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ into a single matrix gives:

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix}$$

$X\vec{w} = w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}$

- Then, if $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, linear combinations of $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ can be written as $X\vec{w}$.
- The **span of the columns of** X, or $\text{span}(X)$, consists of all vectors that can be written in the form $X\vec{w}$.

Minimizing projection error in multiple dimensions

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

- **Goal**: Find the vector $\vec{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^T$ such that $\|\vec{e}\| = \|\vec{y} - X\vec{w}\|$ is minimized.
- As we've seen, $\vec{w}$ must be such that:

known

unknown

2 equations with 2 variables w_1, w_2

$$\left\{ \begin{aligned} \vec{x}^{(1)} \cdot \left(\vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}\right) &= 0 \\ \vec{x}^{(2)} \cdot \underbrace{\left(\vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}\right)}_{\vec{e}} &= 0 \end{aligned} \right.$$

- How can we use our knowledge of matrices to rewrite this system of equations as a single equation?

Simplifying the system of equations, using matrices

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

$$\vec{x}^{(1)} \cdot \left(\vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}\right) = 0$$

$$\vec{x}^{(2)} \cdot \underbrace{\left(\vec{y} - w_1\vec{x}^{(1)} - w_2\vec{x}^{(2)}\right)}_{\vec{e}} = 0$$

$$w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)} = X\vec{w}$$

$$\vec{e} = \vec{y} - X\vec{w}$$

plug in

$$\vec{x}^{(1)} (\vec{y} - X\vec{w}) = 0$$

$$\vec{x}^{(2)} (\vec{y} - X\vec{w}) = 0$$

Simplifying the system of equations, using matrices

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

1. $w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}$ can be written as $X\vec{w}$, so $\vec{e} = \vec{y} - X\vec{w}$.
2. The condition that $\vec{e}$ must be orthogonal to each column of X is equivalent to condition that $X^T\vec{e} = 0$.

$\downarrow$

$\vec{y} - X\vec{w}$

$$\begin{cases} \boxed{\vec{x}^{(1)} \cdot (\vec{y} - X\vec{w})} = 0 \\ \boxed{\vec{x}^{(2)} \cdot (\vec{y} - X\vec{w})} = 0 \end{cases}$$

$$\underbrace{\vec{y} - X\vec{w}}_{\vec{e}}$$

$\Downarrow$ combine into a single matvec eq

$$\boxed{X^T \vec{e} = X^T(\vec{y} - X\vec{w}) = 0}$$

$$X = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix}$$

$$X^T = \begin{bmatrix} - & \vec{x}^{(1)} & - \\ - & \vec{x}^{(2)} & - \end{bmatrix}$$

$$X^T \cdot \vec{e} = \underbrace{\begin{bmatrix} - & \vec{x}^{(1)} & - \\ - & \vec{x}^{(2)} & - \end{bmatrix}}_{X^T} \vec{e} = \begin{bmatrix} \boxed{\vec{x}^{(1)} \cdot \vec{e}} \\ \boxed{\vec{x}^{(2)} \cdot \vec{e}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}$$

row perspective

The normal equations

$$\vec{X} = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

- **Goal**: Find the vector $\vec{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^T$ such that $\|\vec{e}\| = \|\vec{y} - X\vec{w}\|$ is minimized.
- We now know that it is the vector $\vec{w}^*$ such that:

$$\begin{aligned} X^T\vec{e} &= 0 \\ X^T(\vec{y} - X\vec{w}^*) &= 0 \\ X^T\vec{y} - X^TX\vec{w}^* &= 0 \\ \implies X^TX\vec{w}^* &= X^T\vec{y} \end{aligned}$$

$$\vec{w}^* = \underbrace{(X^TX)^{-1}}_{\text{if invertible}} X^T\vec{y}$$

- The last statement is referred to as the **normal equations**.

The general solution to the normal equations

$$X \in \mathbb{R}^{n \times d} \qquad \vec{y} \in \mathbb{R}^n$$

- **Goal, in general**: Find the vector $\vec{w} \in \mathbb{R}^d$ such that $\|\vec{e}\| = \|\vec{y} - X\vec{w}\|$ is minimized.
- We now know that it is the vector $\vec{w}^*$ such that:

$$X^T\vec{e} = 0$$
$$\implies X^TX\vec{w}^* = X^T\vec{y}$$

- Assuming X^TX is invertible, this is the vector:

$$\boxed{\vec{w}^* = (X^TX)^{-1}X^T\vec{y}}$$

  - This is a big assumption, because it requires X^TX to be **full rank**.
  - If X^TX is not full rank, then there are infinitely many solutions to the normal equations, $X^TX\vec{w}^* = X^T\vec{y}$.

What does it mean?

- **Original question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
- **Final answer**: It is the vector $X\vec{w}^*$, where:

$$\vec{w}^* = (X^TX)^{-1}X^T\vec{y}$$

- Revisiting our example:

$$X = \begin{bmatrix} | & | \\ \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & 0 \\ 3 & 4 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

- Using a computer gives us $\vec{w}^* = (X^TX)^{-1}X^T\vec{y} \approx \begin{bmatrix} 0.7289 \\ 1.6300 \end{bmatrix}$.
- So, the vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ closest to $\vec{y}$ is $0.7289\vec{x}^{(1)} + 1.6300\vec{x}^{(2)}$.

An optimization problem, solved

- We just used linear algebra to solve an **optimization problem**.
- Specifically, the function we minimized is:

$$\text{error}(\vec{w}) = \|\vec{y} - X\vec{w}\|$$

  - This is a function whose input is a vector, $\vec{w}$, and whose output is a scalar!
- The input, $\vec{w}^*$, to $\text{error}(\vec{w})$ that minimizes it is:

$$\vec{w}^* = (X^TX)^{-1}X^T\vec{y}$$

- We're going to use this frequently!