

Lecture 7

Orthogonal Projections

DSC 40A, Summer 2024

Announcements

- Homework 3 is due on **Friday, August 16th**. *Submit in groups of 2-4.*
- Discussion is today, right after class. Groupwork is due tonight at 11:59p.
- The Midterm Exam is on **Thursday, August 22nd** in class.
 - Lots of review material on practice.dsc40a.com.
 - Many past exams, and collections of practice questions organized by topic.

Agenda

- Spans and projections.
- Matrices.
- Spans and projections, revisited.
- Regression and linear algebra.

Question 🤔

Answer at q.dsc40a.com

Remember, you can always ask questions at [q.dsc40a.com!](http://q.dsc40a.com)

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

Spans and projections

Orthogonal projection

- Question: What vector in $\text{span}(\vec{x})$ is closest to \vec{y} ?

- Answer: It is the vector $w^* \vec{x}$, where:

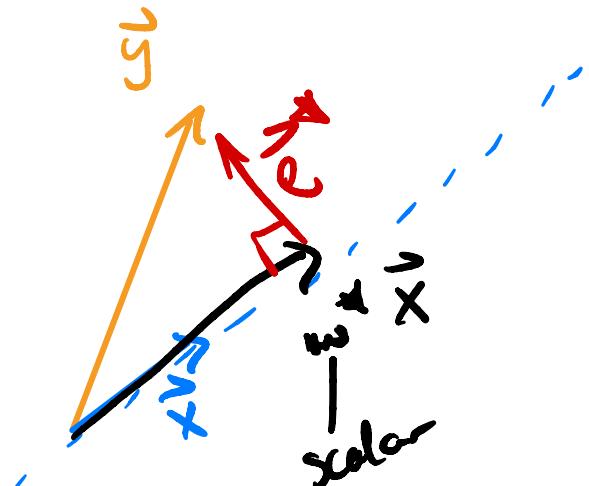
$$w^* = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}}$$

$$\vec{e} \cdot w\vec{x} = 0$$
$$(\vec{y} - w\vec{x}) \cdot (w\vec{x}) = 0$$

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

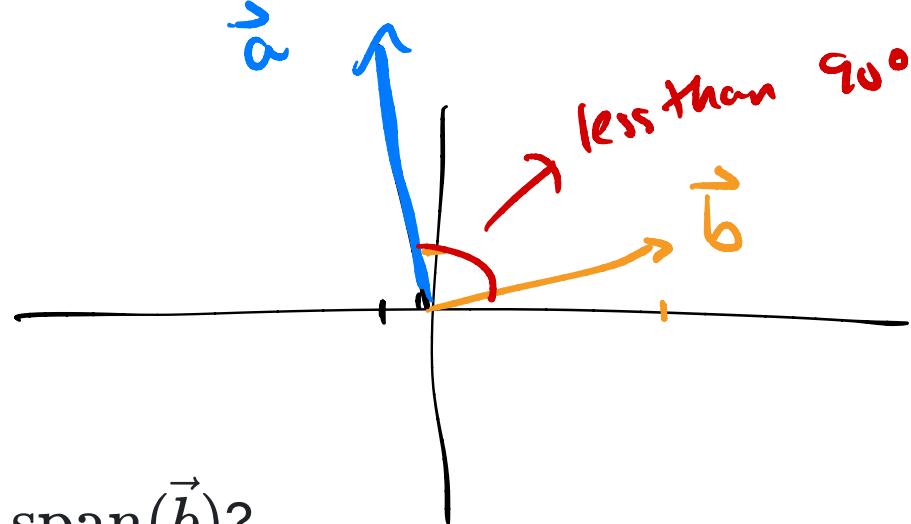
$$\text{error}(w) = \|\vec{e}\| = \|\vec{y} - w\vec{x}\|$$

- 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.

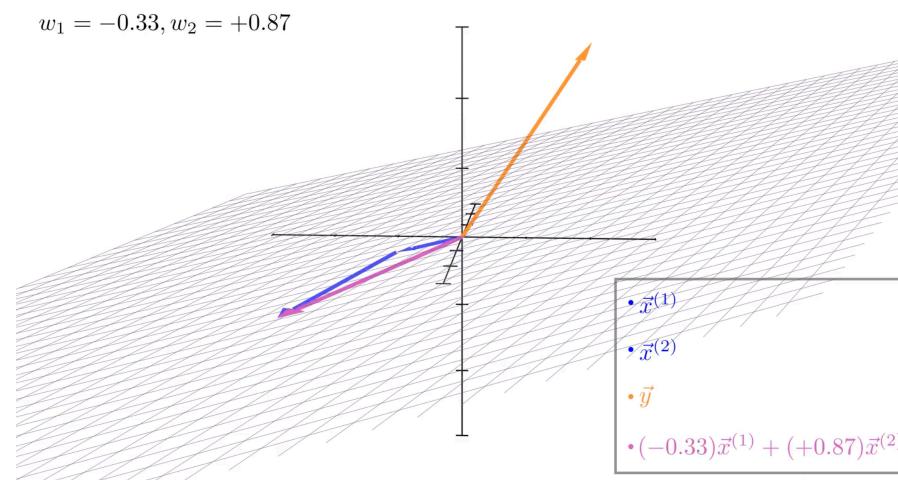
$$w^* = \frac{\vec{b} \cdot \vec{a}}{\vec{b} \cdot \vec{b}} = \frac{5(-1) + 2(9)}{(-1)^2 + (9)^2} = \frac{-5 + 18}{1 + 81} = \frac{13}{82}$$

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

$$\boxed{\frac{13}{82} \cdot \vec{b}}$$

Moving to multiple dimensions

- Let's now consider three vectors, \vec{y} , $\vec{x}^{(1)}$, and $\vec{x}^{(2)}$, all in \mathbb{R}^n .
not exponents
- Question:** What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to \vec{y} ?
→ *2 vectors in \mathbb{R}^3
span at most
2 dimensional
were 3D*
- 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 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:

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

\vec{e}

perpendicular
to
this span

Minimizing projection error in multiple dimensions

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

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

- 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:

$$\begin{aligned} (\underbrace{w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}}_{\vec{e}}) \cdot \vec{e} &= (w_1 \vec{x}^{(1)}) \cdot \vec{e} + (w_2 \vec{x}^{(2)}) \cdot \vec{e} \\ &= w_1 (\cancel{\vec{x}^{(1)} \cdot \vec{e}}) + w_2 (\cancel{\vec{x}^{(2)} \cdot \vec{e}}) \\ &= 0! \end{aligned}$$

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

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

- 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:

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

- 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.

Scalars : \times Matrices : X
vectors : \vec{x}

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

2×3 → set of all matrices with 2 rows and 3 columns

- 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.

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad \begin{bmatrix} 5 \\ 5 \end{bmatrix} \quad \begin{bmatrix} 8 \\ -3 \end{bmatrix}$$

Matrix addition and scalar multiplication

like vectors

- We can add two matrices only if they have the same dimensions.
- Addition occurs elementwise:

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

- Scalar multiplication occurs elementwise, too:

$$2 \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} = \begin{bmatrix} 4 & 10 & 16 \\ -2 & 10 & -6 \end{bmatrix}$$

Matrix-matrix multiplication *order matter!*

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

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

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

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

$$A^T A = \begin{bmatrix} 2(2) + (-1)(-1) = 5 \\ ? \\ 8(2) + (-3)(-1) = 19 \end{bmatrix} \quad \begin{bmatrix} 2(5) + (-1)(5) = 5 \\ ? \\ 8(5) + (-3)(5) = 25 \end{bmatrix}$$

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. $A(BC) \equiv (AB)C$.
- C. $AB = BA$.
- D. $(A + B)^T = A^T + B^T$.
- E. $(AB)^T = B^T A^T$.

$$A_{5 \times 7} B_{7 \times 5} = 5 \times 5$$

↓ different dimensions!

$$B_{7 \times 5} A_{5 \times 7} = 7 \times 7$$

Matrix-vector multiplication

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

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

- 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?

$$A_{n \times d} \vec{v}_{d \times 1} = (A\vec{v})_{n \times 1} \quad \vec{v} \in \mathbb{R}^d, d \text{ components}$$

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

$$(\vec{v}^T)_{1 \times n} A_{n \times d} \Rightarrow \vec{v} \in \mathbb{R}^n, n \text{ components}$$

One view of matrix-vector multiplication

- 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 $A\vec{v}$?

$$\begin{array}{r} 4 \\ 2 \cdot 2 + 5 \cdot -1 + 8 \cdot -5 = -41 \\ -2 - 5 + 15 = 8 \end{array}$$

"across, then down"

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

$\vec{v} \in \mathbb{R}^3$
 $A\vec{v} \in \mathbb{R}^{2 \times 1}$

$$A\vec{v} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}$$

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 $A\vec{v}$?

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 2 \\ -1 \\ -5 \end{bmatrix}$$
$$A\vec{v} = 2 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + (-1) \begin{bmatrix} 5 \\ 5 \end{bmatrix} + (-5) \begin{bmatrix} 8 \\ -3 \end{bmatrix} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}$$

linear combination of $\underbrace{\left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 8 \\ -3 \end{bmatrix} \right\}}_{\text{columns of } A}$

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} .

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

\downarrow

$n \times d$ $d \times 1$

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

Result \in
 $\mathbb{R}^n \times 1$

Spans and projections, revisited

$$\omega_1 \vec{x}^{(1)} + \omega_2 \vec{x}^{(2)}$$

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

Answer : w_1, w_2 such that:

$$\begin{aligned} \vec{x}^{(1)} \cdot \vec{e} &= 0 \\ \vec{x}^{(2)} \cdot \vec{e} &= 0 \end{aligned} \quad \left. \begin{array}{l} 2 \text{ equations} \\ 2 \text{ unknowns} \end{array} \right.$$

Matrix-vector products create linear combinations of columns! *Same!*

$$\vec{x}^{(1)} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}$$

$$\vec{x}^{(2)} = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

$$\frac{w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}}{\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}$$

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

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

$$X\vec{w} = w_1 \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} + w_2 \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}$$

Same!

- 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

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

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

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

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

- 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

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

$$\begin{aligned} X\vec{w} &= w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)} \\ \Rightarrow \vec{e} &= \vec{y} - w_1 \vec{x}^{(1)} - w_2 \vec{x}^{(2)} \\ &= \vec{y} - X\vec{w} \end{aligned}$$

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

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

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

Simplifying the system of equations, using matrices

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

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 $X^T \vec{e} = \vec{0}$.

↳ rows of X^T are columns of X

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

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

↓ combine

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

Solve me for \vec{w} !!

under our control

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

Project columns of X onto X^T

$$X^T = \begin{bmatrix} \vec{x}^{(1)T} \\ \vec{x}^{(2)T} \end{bmatrix}_{2 \times 3}$$

$$= \begin{bmatrix} 2 & 5 & 3 \\ 0 & 4 \end{bmatrix}$$

example \vec{e}

$$\begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix}$$

$$X^T \vec{e} = \begin{bmatrix} \vec{x}^{(1)T} \\ \vec{x}^{(2)T} \end{bmatrix} \vec{e} = \begin{bmatrix} \vec{x}^{(1)T} \vec{e} \\ \vec{x}^{(2)T} \vec{e} \end{bmatrix} = \vec{0}$$

$\hookrightarrow \vec{a}^T \vec{b} = \vec{a} \cdot \vec{b}$

$$x^v = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$x^v ()$$

The normal equations

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

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

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

last slide

system of linear equations
many equations

Orthogonal

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

The general solution to the normal equations

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

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

$$\begin{aligned} \mathbf{X}^T \vec{e} &= 0 \\ \implies \mathbf{X}^T \mathbf{X} \vec{w}^* &= \mathbf{X}^T \vec{\mathbf{y}} \end{aligned}$$

- Assuming $\mathbf{X}^T \mathbf{X}$ is invertible, this is the vector:

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

all columns of $\mathbf{X}^T \mathbf{X}$
are linearly independent

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

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 $\mathbf{X}\vec{w}^*$, where:

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

- Revisiting our example:

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

- Using a computer gives us $\vec{w}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y} \approx \begin{bmatrix} 0.7289 \\ 1.6300 \end{bmatrix}$.
- 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:

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

- 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:

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

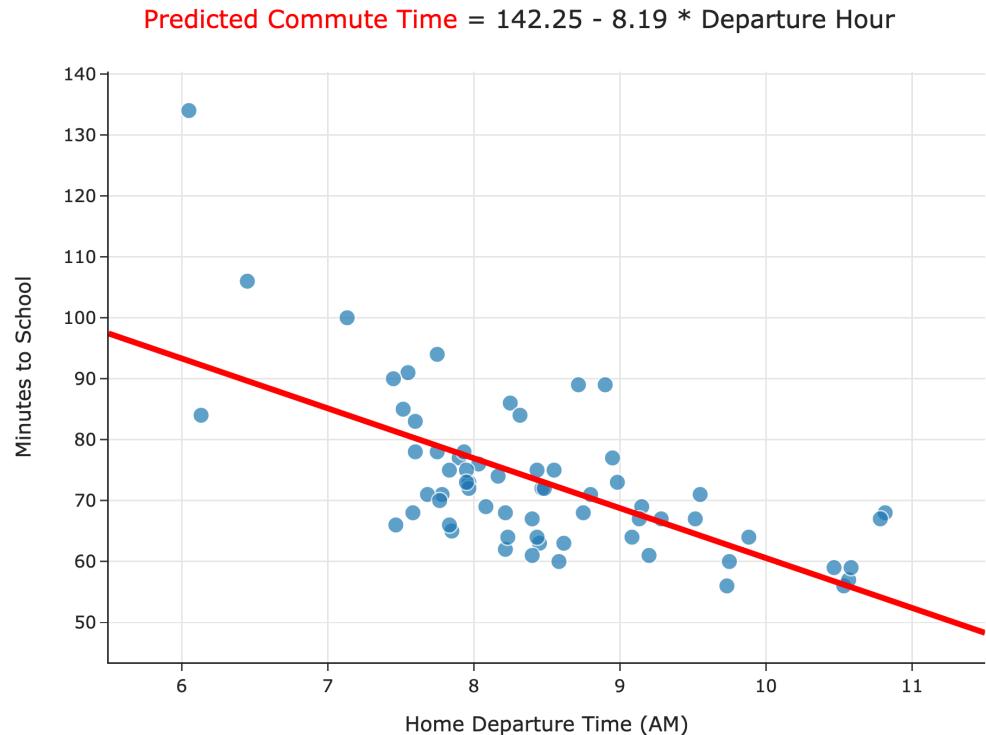
- We're going to use this frequently!

Regression and linear algebra

Wait... why do we need linear algebra?

- Soon, we'll want to make predictions using more than one feature.
 - Example: Predicting commute times using departure hour and temperature.
- Thinking about linear regression in terms of **matrices and vectors** will allow us to find hypothesis functions that:
 - Use multiple features (input variables).
 - Are non-linear in the features, e.g. $H(x) = w_0 + w_1x + w_2x^2$.
- Let's see if we can put what we've just learned to use.

Simple linear regression, revisited



- **Model:** $H(x) = w_0 + w_1x$.
- **Loss function:** $(y_i - H(x_i))^2$.
- To find w_0^* and w_1^* , we minimized empirical risk, i.e. average loss:

$$R_{\text{sq}}(H) = \frac{1}{n} \sum_{i=1}^n (y_i - H(x_i))^2$$

- **Observation:** $R_{\text{sq}}(w_0, w_1)$ kind of looks like the formula for the norm of a vector,

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}.$$

Regression and linear algebra

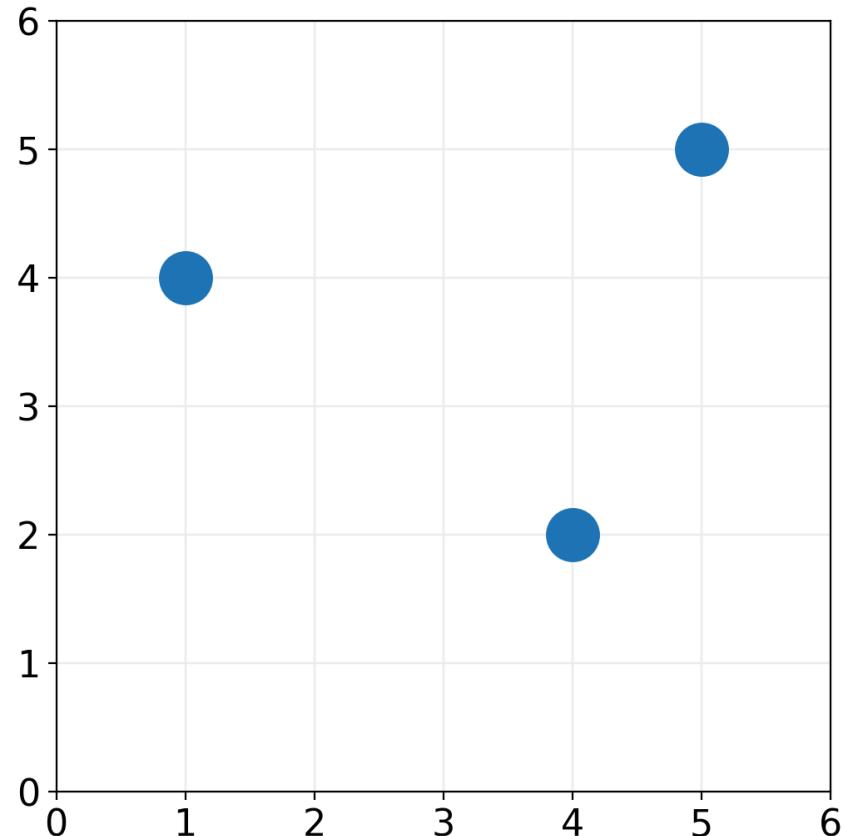
Let's define a few new terms:

- The **observation vector** is the vector $\vec{y} \in \mathbb{R}^n$. This is the vector of observed "actual values".
- The **hypothesis vector** is the vector $\vec{h} \in \mathbb{R}^n$ with components $H(x_i)$. This is the vector of predicted values.
- The **error vector** is the vector $\vec{e} \in \mathbb{R}^n$ with components:

$$e_i = y_i - H(x_i)$$

Example

Consider $H(x) = 2 + \frac{1}{2}x$.



$$\vec{y} =$$

$$\vec{h} =$$

$$\vec{e} = \vec{y} - \vec{h} =$$

$$R_{\text{sq}}(\mathcal{H}) = \frac{1}{n} \sum_{i=1}^n (\mathcal{y}_i - H(x_i))^2$$

$$=$$

Regression and linear algebra

Let's define a few new terms:

- The **observation vector** is the vector $\vec{y} \in \mathbb{R}^n$. This is the vector of observed "actual values".
- The **hypothesis vector** is the vector $\vec{h} \in \mathbb{R}^n$ with components $H(x_i)$. This is the vector of predicted values.
- The **error vector** is the vector $\vec{e} \in \mathbb{R}^n$ with components:

$$e_i = y_i - H(x_i)$$

- **Key idea:** We can rewrite the mean squared error of H as:

$$R_{\text{sq}}(H) = \frac{1}{n} \sum_{i=1}^n (y_i - H(x_i))^2 = \frac{1}{n} \|\vec{e}\|^2 = \frac{1}{n} \|\vec{y} - \vec{h}\|^2$$

The hypothesis vector

- The **hypothesis vector** is the vector $\vec{h} \in \mathbb{R}^n$ with components $H(x_i)$. This is the vector of predicted values.
- For the linear hypothesis function $H(x) = w_0 + w_1 x$, the hypothesis vector can be written:

$$\vec{h} = \begin{bmatrix} w_0 + w_1 x_1 \\ w_0 + w_1 x_2 \\ \vdots \\ w_0 + w_1 x_n \end{bmatrix} =$$

Rewriting the mean squared error

- Define the **design matrix** $\mathbf{X} \in \mathbb{R}^{n \times 2}$ as:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

- Define the **parameter vector** $\vec{w} \in \mathbb{R}^2$ to be $\vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$.
- Then, $\vec{h} = \mathbf{X}\vec{w}$, so the mean squared error becomes:

$$R_{\text{sq}}(\mathbf{H}) = \frac{1}{n} \|\vec{y} - \vec{h}\|^2 \implies R_{\text{sq}}(\vec{w}) = \frac{1}{n} \|\vec{y} - \mathbf{X}\vec{w}\|^2$$

What's next?

- To find the optimal model parameters for simple linear regression, w_0^* and w_1^* , we previously minimized:

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^n (\textcolor{orange}{y}_i - (w_0 + w_1 \textcolor{blue}{x}_i))^2$$

- Now that we've reframed the simple linear regression problem in terms of linear algebra, we can find w_0^* and w_1^* by minimizing:

$$R_{\text{sq}}(\vec{w}) = \frac{1}{n} \|\vec{\textcolor{orange}{y}} - \textcolor{blue}{X} \vec{w}\|^2$$

- We've already solved this problem! Assuming $\textcolor{blue}{X}^T \textcolor{blue}{X}$ is invertible, the best \vec{w} is:

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