Lecture 6

Dot Products and Projections

DSC 40A, Summer 2024

Announcements

- Homework 2 is due **tonight**. Remember that using the Overleaf template is required for Homework 2 (and only Homework 2).
- Check out the new FAQs page and the tutor-created supplemental resources on the course website.
    - The proof that we were going to cover last class (that $R_{\text{sq}}(w_0^*, w_1^*) = \sigma_y^2(1 - r^2)$) is now in the FAQs page, under Week 3.

Agenda

- Why?
- Dot products.
- Spans and projections.

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.

Dot products

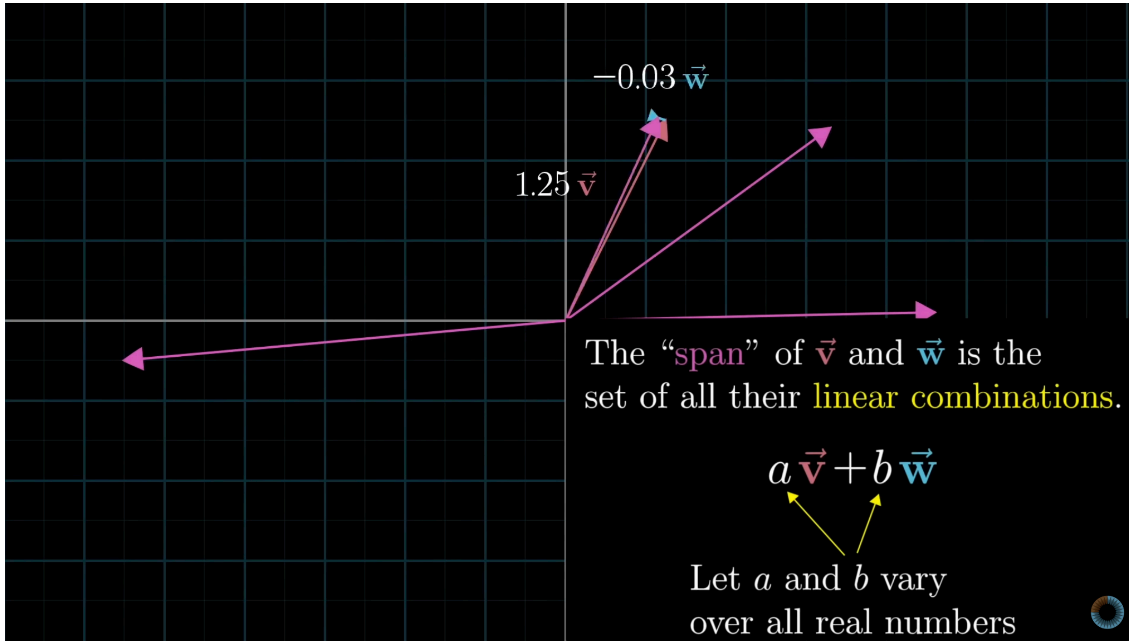
Wait... why do we need linear algebra?

- 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). x_1, x_2

    $+ w_3 x_2$
  - Are non-linear in the features, e.g. $H(x) = w_0 + w_1 x_1 + w_2 x_1^2$.
- Before we dive in, let's review.

Spans of vectors

- One of the most important ideas you'll need to remember from linear algebra is the concept of the **span** of one or more vectors.
- To jump start our review of linear algebra, let's start by watching 🎥 **this video by 3blue1brown**.



Warning ⚠️

- We're **not** going to cover every single detail from your linear algebra course.
- There will be facts that you're expected to remember that we won't explicitly say.
    - For example, if A and B are two matrices, then $AB \neq BA$.
    - This is the kind of fact that we will only mention explicitly if it's directly relevant to what we're studying.
    - But you still need to know it, and it may come up in homework questions.
- We **will** review the topics that you really need to know well.

in overleaf: \mathbb{R}

Vectors

$\mathbb{R}$: real numbers

→ there are n reals in our vector

- A **vector** in $\mathbb{R}^n$ is an **ordered collection of n numbers**.
- We use lower-case letters with an arrow on top to represent vectors, and we usually write vectors as **columns**.

\vec{x}

$$\vec{v} = \begin{bmatrix} 8 \\ 3 \\ -2 \\ 5 \end{bmatrix}$$

- Another way of writing the above vector is $\vec{v} = [8, 3, -2, 5]^T$.

transpose

- Since $\vec{v}$ has four **components**, we say $\vec{v} \in \mathbb{R}^4$.

"elements"

"in"

$$\vec{v} = \begin{array}{c} \begin{bmatrix} 5 \\ 3 \end{bmatrix} \end{array} \begin{array}{c} x \\ y \end{array}$$

$$\|\vec{v}\| = \sqrt{5^2+3^2} = \sqrt{34}$$

The geometric interpretation of a vector

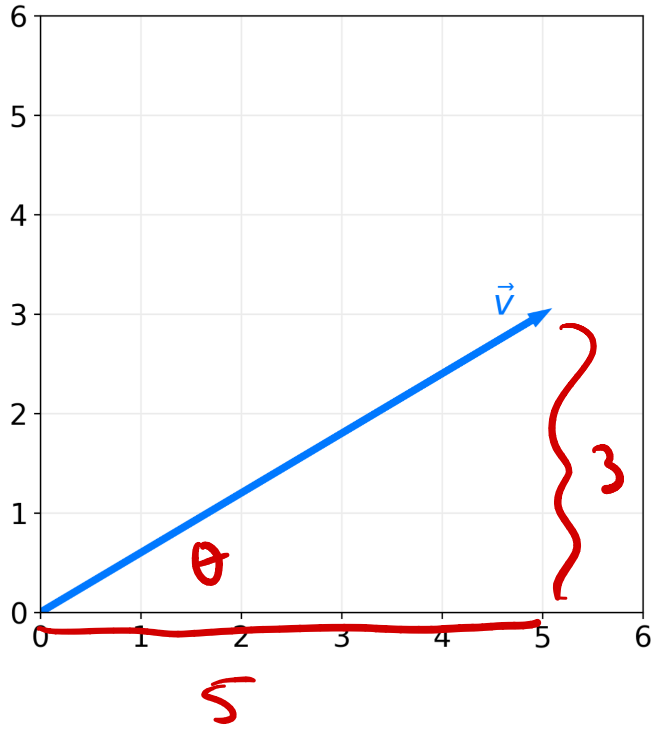
- A vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ is an arrow to the point $(v_1, v_2, \ldots, v_n)$ from the origin.

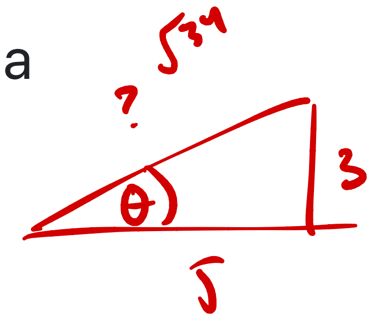
- The **length**, or L_2 **norm**, of $\vec{v}$ is:

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

multidimensional pythagorean theorem

- A vector is sometimes described as an object with a **magnitude/length** and **direction**.





$$\tan\theta = \frac{3}{5}$$

$$\theta = \tan^{-1}\left(\frac{3}{5}\right)$$

Dot product: coordinate definition

same # elements!
same dimension!

- The **dot product** of two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$ is written as:

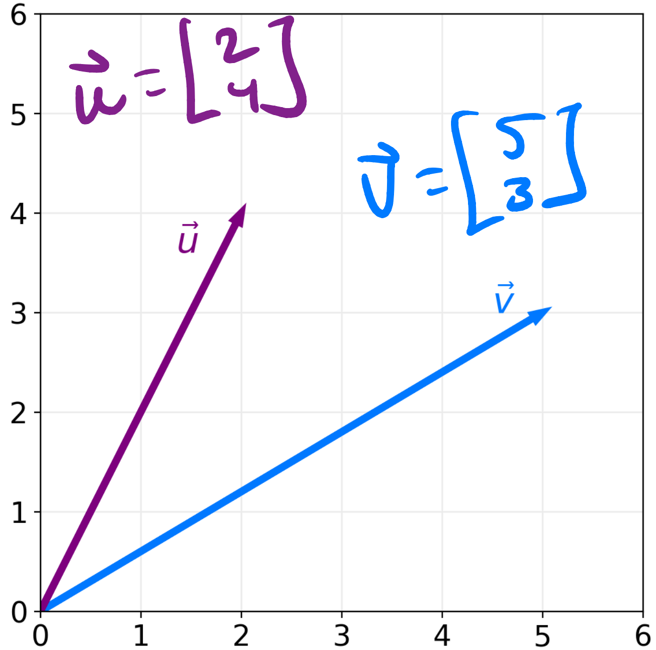
$\mathbb{R}^n \times \mathbb{R}^n \Rightarrow \mathbb{R}$

$$\vec{u} \cdot \vec{v} = \vec{u}^\top \vec{v}$$

- The computational definition of the dot product:

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

- The result is a **scalar**, i.e. a single number.



$\vec{u} \cdot \vec{v} = (2)(5) + (4)(3) = 10 + 12 = \boxed{22}$ scalar!

$\vec{u}^T \vec{v} = \begin{bmatrix} 2 & 4 \end{bmatrix}_{1 \times 2} \begin{bmatrix} 5 \\ 3 \end{bmatrix}_{2 \times 1}$

match

Question 🤔

Answer at q.dsc40a.com

Which of these is another expression for the length of $\vec{v}$?

- ~~A. $\vec{v} \cdot \vec{v}$~~ ?
- ~~B. $\sqrt{\vec{v}^2}$~~
- C. $\sqrt{\vec{v} \cdot \vec{v}}$
- ~~D. $\vec{v}^2$~~
- E. More than one of the above.

$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ n rows, 1 col

$\vec{v} \in \mathbb{R}^n$

$\vec{v}^2$ is undefined!

NO DOT

$\vec{v}_{n \times 1}$ $v_{n \times 1}$

rows # cols don't match

$$\sqrt{\vec{v} \cdot \vec{v}} = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

$$= \|\vec{v}\|$$

$22 = \sqrt{20}\sqrt{34}\cos\theta \Rightarrow \cos\theta = \frac{\cancel{22}\,11}{\cancel{2}\sqrt{5}\cdot\sqrt{34}}$

$\cos\theta = \frac{11}{\sqrt{5}\sqrt{34}}$

Dot product: geometric definition

- The computational definition of the dot product:

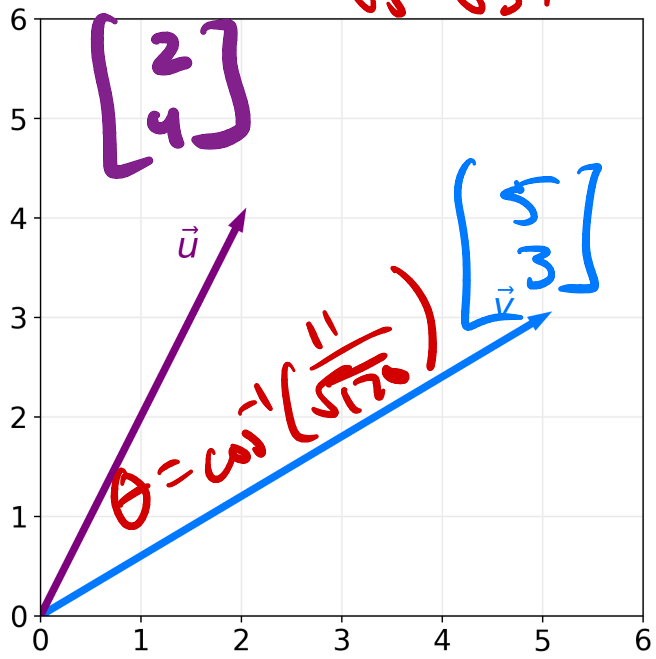
$$\vec{u}\cdot\vec{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

- The geometric definition of the dot product:

$$\vec{u}\cdot\vec{v} = \|\vec{u}\|\|\vec{v}\|\cos\theta$$

where θ is the angle between $\vec{u}$ and $\vec{v}$.

- The two definitions are equivalent! This equivalence allows us to find the angle θ between two vectors.

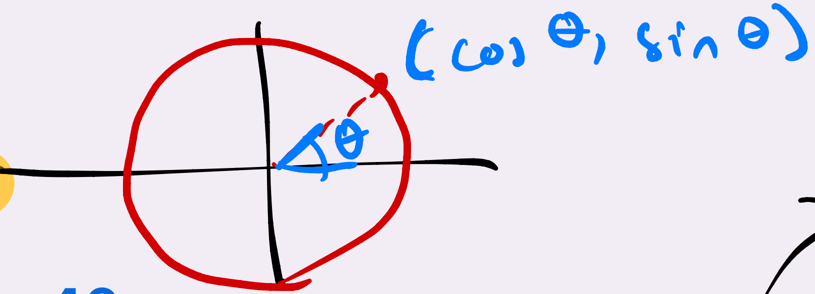


equal!

$\vec{u}\cdot\vec{v} = 2\cdot 5 + 4\cdot 3 = 22$

$= \|\vec{u}\|\cdot\|\vec{v}\|\cdot\cos\theta = \sqrt{2^2+4^2}\cdot\sqrt{5^2+3^2}\cos\theta = \sqrt{20}\sqrt{34}\cos\theta$

Question 🤔



Answer at q.dsc40a.com

What is the value of θ in the plot to the right?

① $\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} = \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ -4 \end{bmatrix} = -12$

||

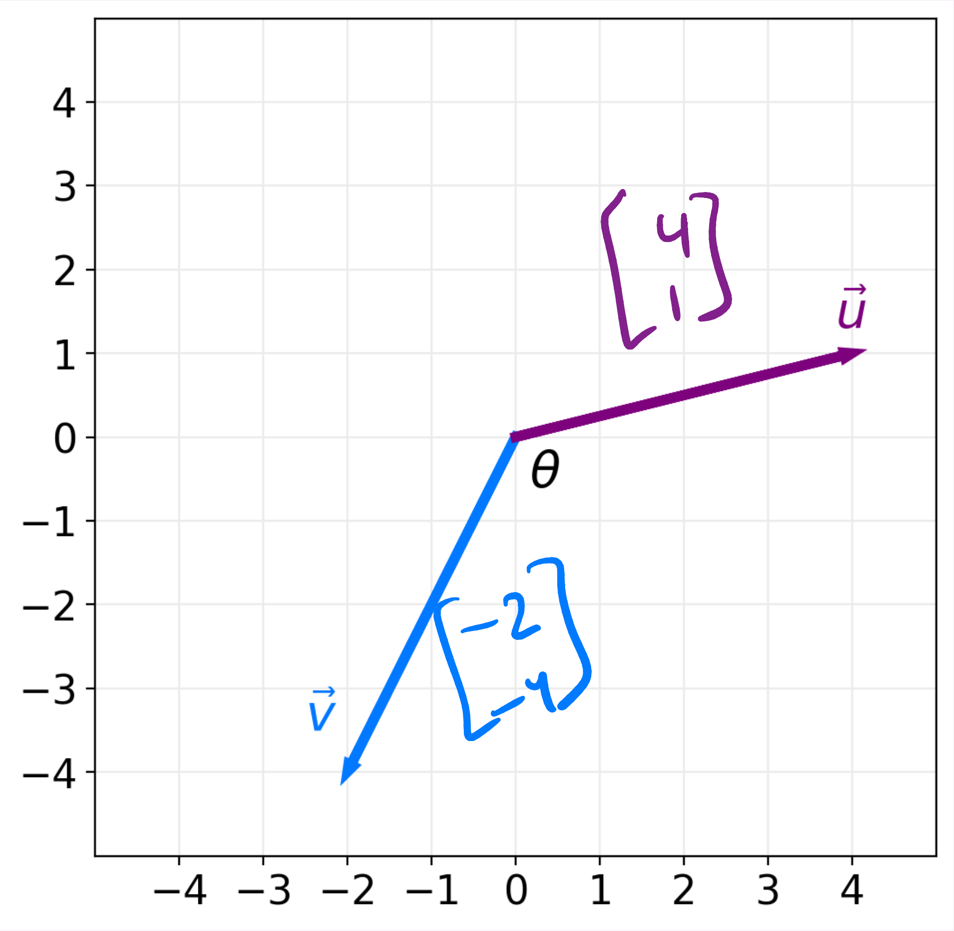
② $\vec{u} \cdot \vec{v} = \|u\| \cdot \|v\| \cos\theta = \sqrt{17}\sqrt{20} \cos\theta$

solve

$-12 = \sqrt{17} \cdot 2\sqrt{5} \cos\theta$

$\Rightarrow \cos\theta = \frac{-\cancel{12}\,6}{\cancel{2}\sqrt{5}\sqrt{17}} = -\frac{6}{\sqrt{85}}$

$\theta = \cos^{-1}\left(\frac{-6}{\sqrt{85}}\right)$



Perpendicular

Orthogonal vectors

"right angle"
"perpendicular"
orthogonal!

- Recall: $\cos 90^{\circ} = 0$.
- Since $\overset{0}{\vec{u} \cdot \vec{v}} = ||\vec{u}||\,||\vec{v}||\,\overset{0}{\cos\theta}$, if the angle between two vectors is 90°, their dot product is $||\vec{u}||\,||\vec{v}||\cos 90^{\circ} = 0$.
- If the angle between two vectors is 90°, we say they are perpendicular, or more generally, **orthogonal**.
- **Key idea**:

$$\boxed{\text{two vectors are } \textbf{orthogonal} \iff \vec{u} \cdot \vec{v} = 0}$$

↳ "if and only if"
bidirectional statement

Exercise

can't be all 0

Find a non-zero vector in $\mathbb{R}^3$ orthogonal to:

$$\vec{v} = \begin{bmatrix} 2 \\ 5 \\ -8 \end{bmatrix}$$

solutions to
$2u_1 + 5u_2 - 8u_3 = 0$

Infinitely many!

$\vec{u} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ $\quad \vec{u}' = \begin{bmatrix} -2 \\ 4 \\ 2 \end{bmatrix} = 2\vec{u}$

$\begin{bmatrix} 0 \\ 8 \\ 5 \end{bmatrix} = \checkmark \; 40 - 40 + 0$

$\vec{u} \cdot \vec{v} = 2(-1) + 5(2) + (-8)(1)$
$= -2 + 10 + (-8)$
$= 0$

$\vec{u}' \cdot \vec{v} = 2(-2) + 5(4) + (-8)(2)$
$= -4 + 20 - 16 = 0$

Spans and projections

$u@v \overset{\text{Python}}{\Rightarrow} \vec{u} \cdot \vec{v}$

$\vec{u} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \Rightarrow -2\vec{u} = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

Adding and scaling vectors

- The sum of two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$ is the **element-wise sum** of their components:

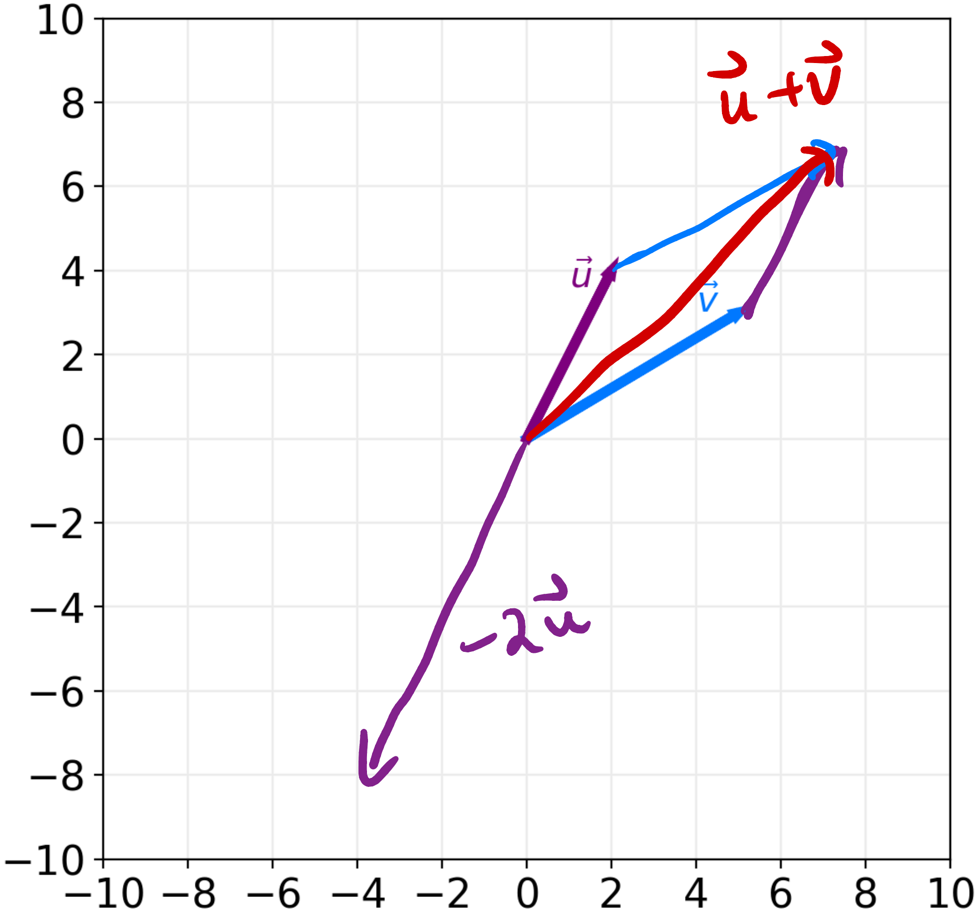
"tip to tail"

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}$$

→ also a vector!

- If c is a scalar, then:

$$c\vec{v} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}$$



Linear combinations

now $\vec{v}_i$ are vectors

- Let $\vec{v}_1, \vec{v}_2, ..., \vec{v}_d$ all be vectors in $\mathbb{R}^n$.
- A **linear combination** of $\vec{v}_1, \vec{v}_2, ..., \vec{v}_d$ is any vector of the form:

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \ldots + a_d\vec{v}_d$$

where $a_1, a_2, ..., a_d$ are all scalars.

0 is OK

"d vectors with n elements each"

$\vec{v}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ $\vec{v}_2 = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$ $\vec{v}_3 = \begin{bmatrix} 0 \\ 9 \end{bmatrix}$

$d=3$, $n=2$

Ex

$2\vec{v}_1 + 1\vec{v}_2 + \frac{1}{9}\vec{v}_3 = \begin{bmatrix} \sim \\ \sim \end{bmatrix}$ a vector in $\mathbb{R}^2$

$0\vec{v}_1 + \vec{v}_2 - \vec{v}_3 = \sim$

⋮

Span

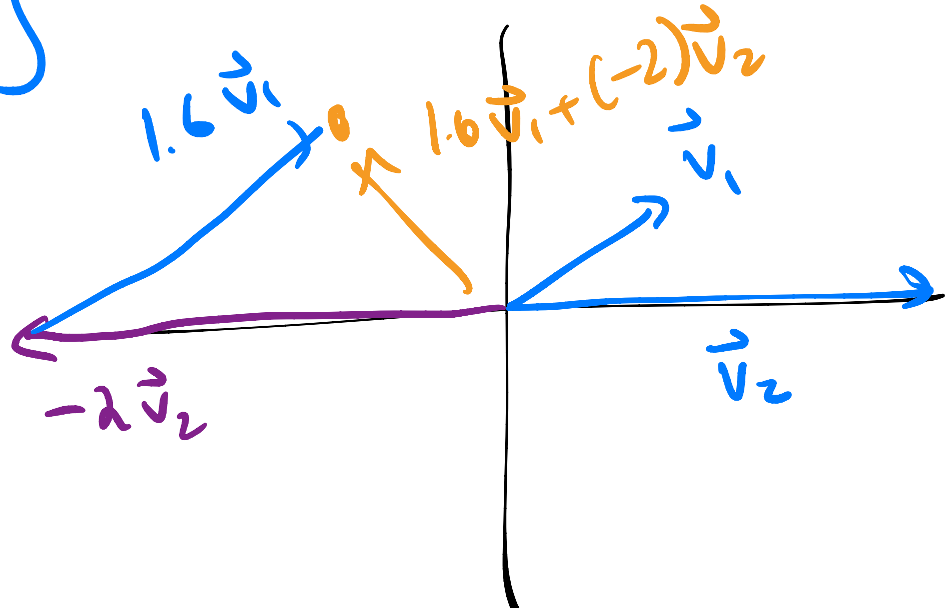
- Let $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_d$ all be vectors in $\mathbb{R}^n$.
- The **span** of $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_d$ is the set of all vectors that can be created using linear combinations of those vectors.
- Formal definition:

$$\text{span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_d) = \{a_1\vec{v}_1 + a_2\vec{v}_2 + \ldots + a_d\vec{v}_d : a_1, a_2, \ldots, a_d \in \mathbb{R}\}$$

Ex

$\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $\vec{v}_2 = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$

$\vec{v}_1$ and $\vec{v}_2$ span all of $\mathbb{R}^2$



Exercise

Let $\vec{v}_1 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$ and let $\vec{v}_2 = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$. Is $\vec{y} = \begin{bmatrix} 5 \\ -5 \end{bmatrix}$ in $\text{span}(\vec{v}_1, \vec{v}_2)$?

Yes! $\vec{v}_1$ and $\vec{v}_2$ are not scalar multiples of each other, point in different directions.

If so, write $\vec{y}$ as a linear combination of $\vec{v}_1$ and $\vec{v}_2$.

$$w_1\vec{v}_1 + w_2\vec{v}_2 = \begin{bmatrix} 5 \\ -5 \end{bmatrix}$$

$$\begin{bmatrix} 2w_1 \\ -3w_1 \end{bmatrix} + \begin{bmatrix} -w_2 \\ 4w_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -5 \end{bmatrix}$$

$$\left.\begin{aligned} 2\overset{3}{w_1} - \overset{1}{w_2} &= 5 \\ -3\overset{3}{w_1} + 4\overset{1}{w_2} &= -5 \end{aligned}\right\} \text{ solve for } w_1, w_2$$

$w_1 = 3$

$w_2 = 1$

To whoever teaches this in the future: the slides afterwards were improved in Lecture 7, so you may want to bring some of that material to this lecture.

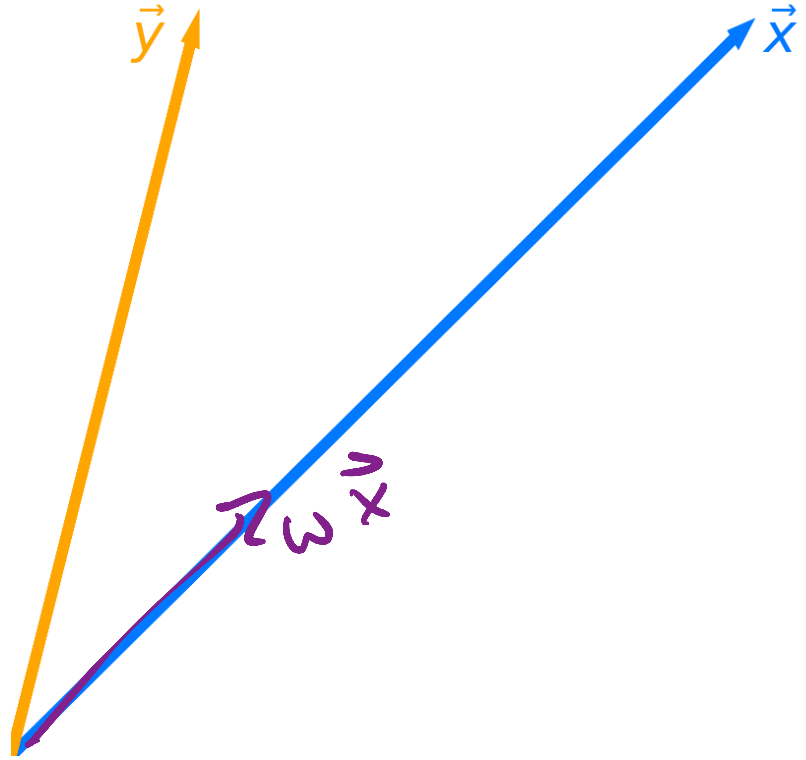
Projecting onto a single vector

- Let $\vec{x}$ and $\vec{y}$ be two vectors in $\mathbb{R}^n$.
- The span of $\vec{x}$ is the set of all vectors of the form:

$$w\vec{x}$$

where $w \in \mathbb{R}$ is a scalar.

- **Question**: What vector in $\text{span}(\vec{x})$ is closest to $\vec{y}$?
- The vector in $\text{span}(\vec{x})$ that is closest to $\vec{y}$ is the **projection of $\vec{y}$ onto** $\text{span}(\vec{x})$.



$w\vec{x} + \vec{e} = \vec{y}$

Projection error

actual predicted / approx

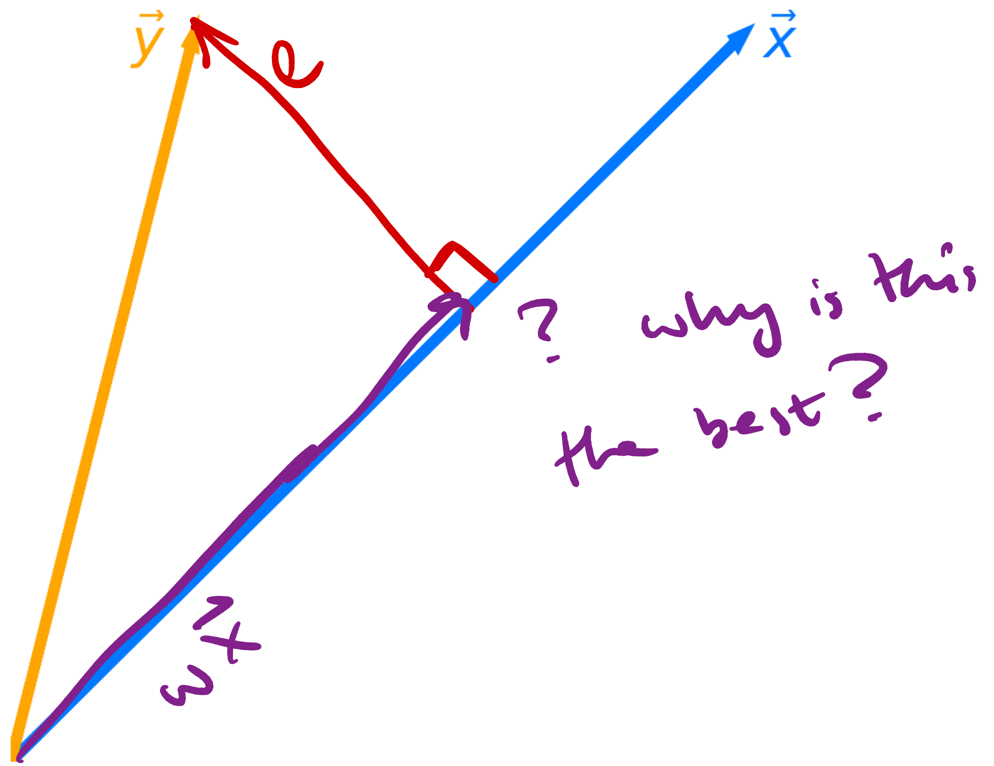
- Let $\vec{e} = \vec{y} - w\vec{x}$ be the **projection error**: that is, the vector that connects $\vec{y}$ to $\text{span}(\vec{x})$.
- **Goal**: **Find the w that makes $\vec{e}$ as short as possible.**
  - That is, minimize:

$$||\vec{e}||$$

  - Equivalently, minimize:

$$||\vec{y} - w\vec{x}||$$

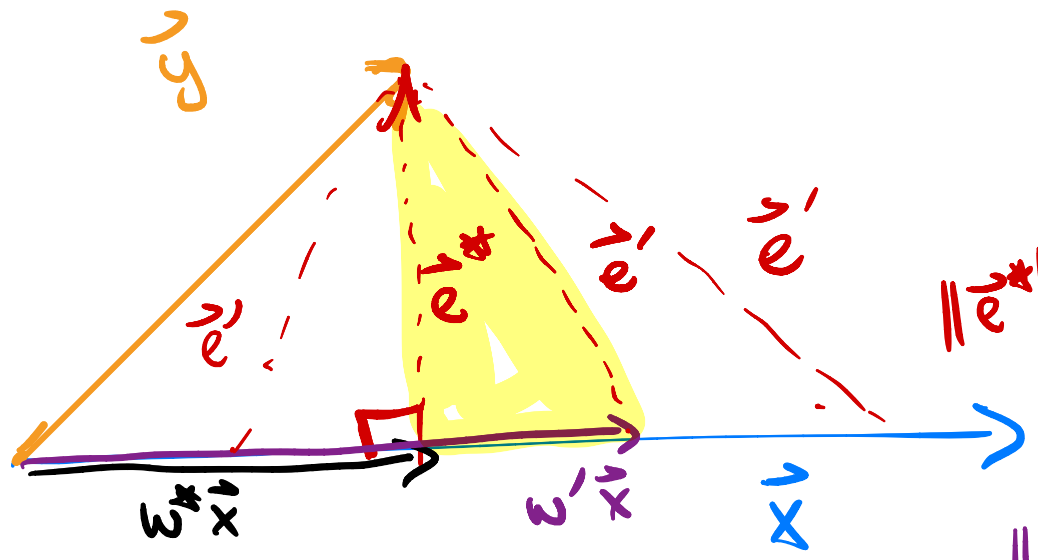
- **Idea**: To make $\vec{e}$ has short as possible, it should be **orthogonal to** $w\vec{x}$.

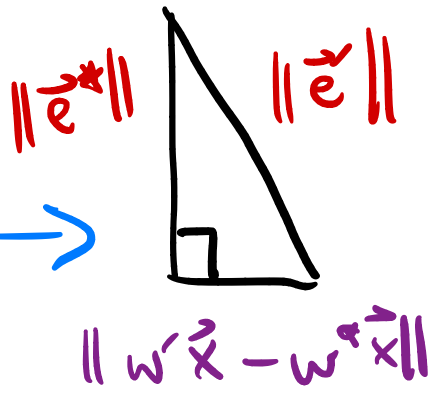


Minimizing projection error

- **Goal: Find the w that makes $\vec{e} = \vec{y} - w\vec{x}$ as short as possible.**
- **Idea**: To make $\vec{e}$ as short as possible, it should be **orthogonal to** $w\vec{x}$.
- Can we prove that making $\vec{e}$ orthogonal to $w\vec{x}$ minimizes $||\vec{e}||$?

Goal: prove that $\vec{e}^*$ is the shortest possible error vector.





$$\|\vec{e}'\|^2 = \|\vec{e}^*\|^2 + \underbrace{\|w'\vec{x} - w^*\vec{x}\|^2}_{0}$$

$\|\vec{e}'\|^2 = \|\vec{e}^*\|^2 +$ something positive

$$\|\vec{e}'\|^2 \geq \|\vec{e}^*\|^2$$

$\vec{e}^*$ is the shortest possible error vector

Minimizing projection error

- **Goal: Find the w that makes $\vec{e} = \vec{y} - w\vec{x}$ as short as possible.**
- Now we know that to minimize $||\vec{e}||$, $\vec{e}$ must be orthogonal to $w\vec{x}$.
- Given this fact, how can we solve for w?

$\vec{e}$ orthogonal to $w\vec{x}$ $\Rightarrow$ $\boxed{w\vec{x} \cdot \vec{e} = 0}$

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

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

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

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

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

divide

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

the w that makes the error vector as short as possible!!

Orthogonal projection

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

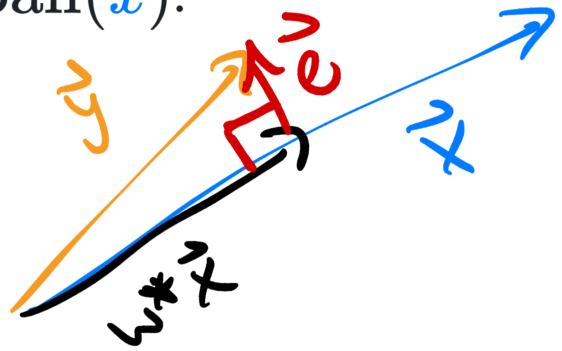
$$w^* = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}}$$

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

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

$\vec{e}$ orthogonal to $\vec{x}$, smallest $\|\vec{e}\|$

- 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

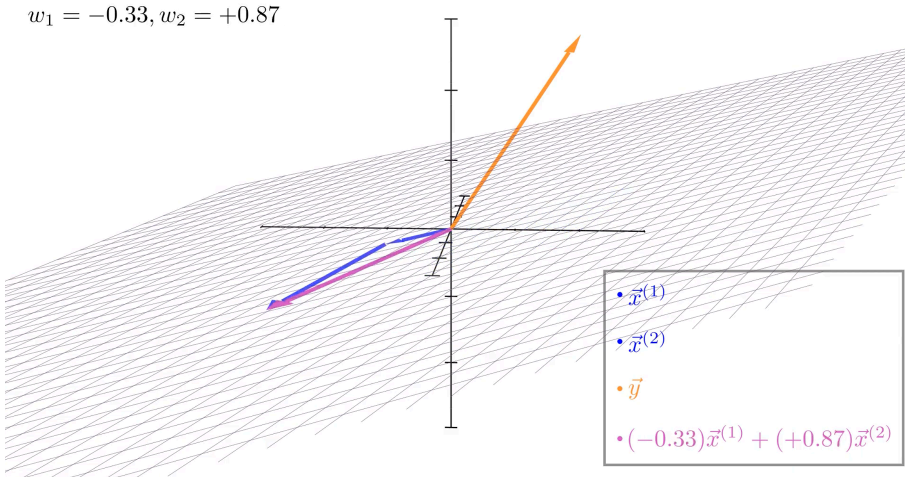
Let $\vec{a} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 9 \end{bmatrix}$.

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.

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 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:

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

What's next?

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

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

- **Solution**: Combine $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ into a single **matrix**, X, and express $w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}$ as a **matrix-vector multiplication**, $X\vec{w}$.

- **Next time**: Formulate linear regression in terms of matrices and vectors!