Lecture 6

Dot Products and Projections

DSC 40A, Spring 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.

DSC Undergraduate Town Hall

Monday, April 22nd, 1-3PM

HDSI 123

Come ask questions and voice your feedback about the undergraduate program, while socializing with faculty! 🧑‍🏫

🧑‍🎓 Your favorite professors will be there – and so will free cookies! 🍪🥠

Scan me to RSVP!

Agenda

- Recap: Friends of simple linear regression.
- Dot products.
- Spans and projections.

Question 🤔

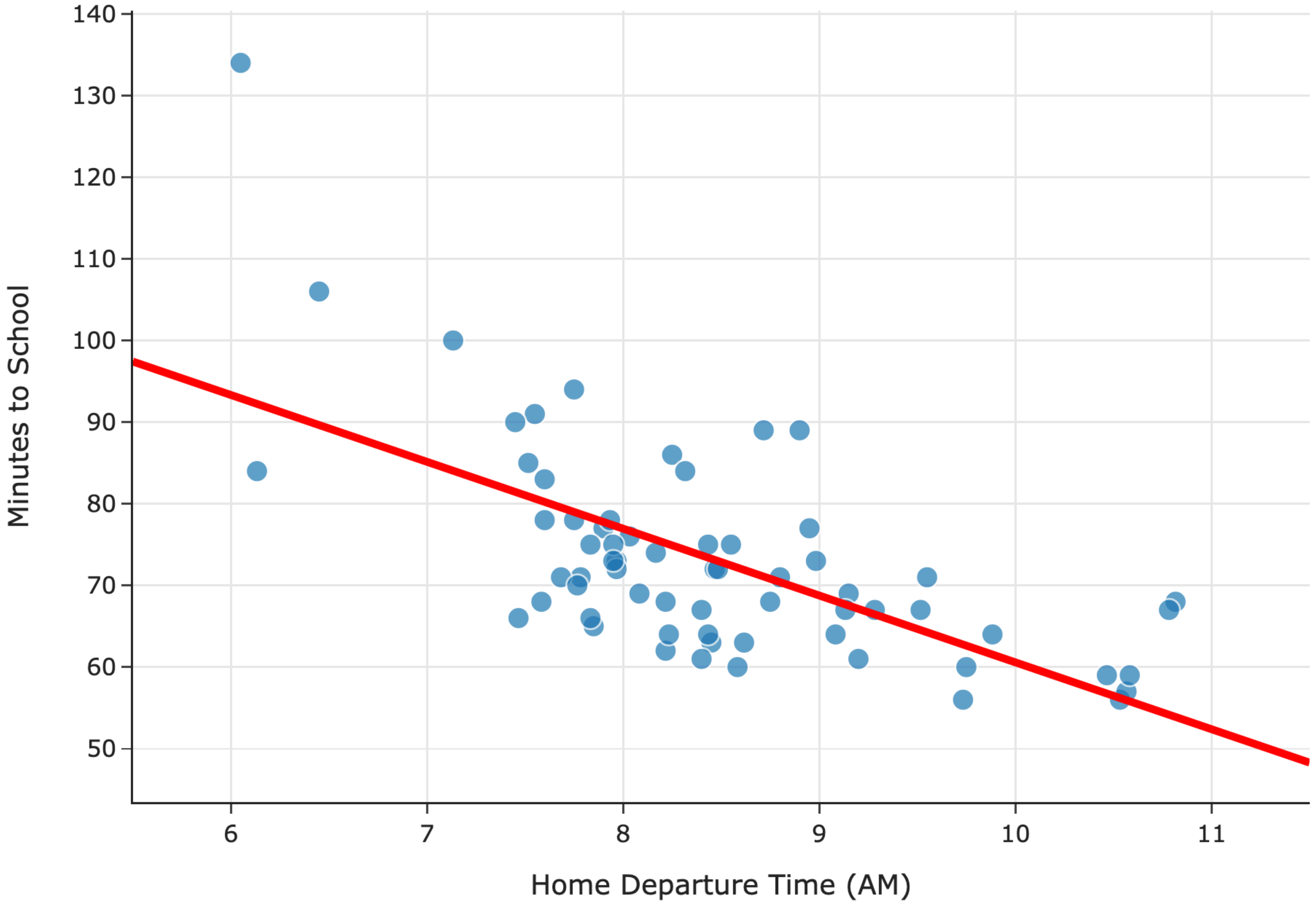
Answer at q.dsc40a.com

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.

Recap: Friends of simple linear regression

Predicted Commute Time = 142.25 - 8.19 * Departure Hour



Simple linear regression

- Model: $H(x) = w_0 + w_1 x$.
- Loss function: squared loss, i.e. $L_{\text{sq}}(y_i, H(x_i)) = (y_i - H(x_i))^2$.
- Average loss, i.e. empirical risk:

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (w_0 + w_1 x_i))^2$$

- Optimal model parameters, found by minimizing empirical risk:

$$w_1^* = \frac{\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2} = r \frac{\sigma_y}{\sigma_x} \qquad\qquad w_0^* = \bar{y} - w_1^* \bar{x}$$

Friends of simple linear regression

- Suppose we use squared loss throughout.
- If our model is $H(x) = w_1 x$, it is a **line that is forced through the origin,** $(0, 0)$.

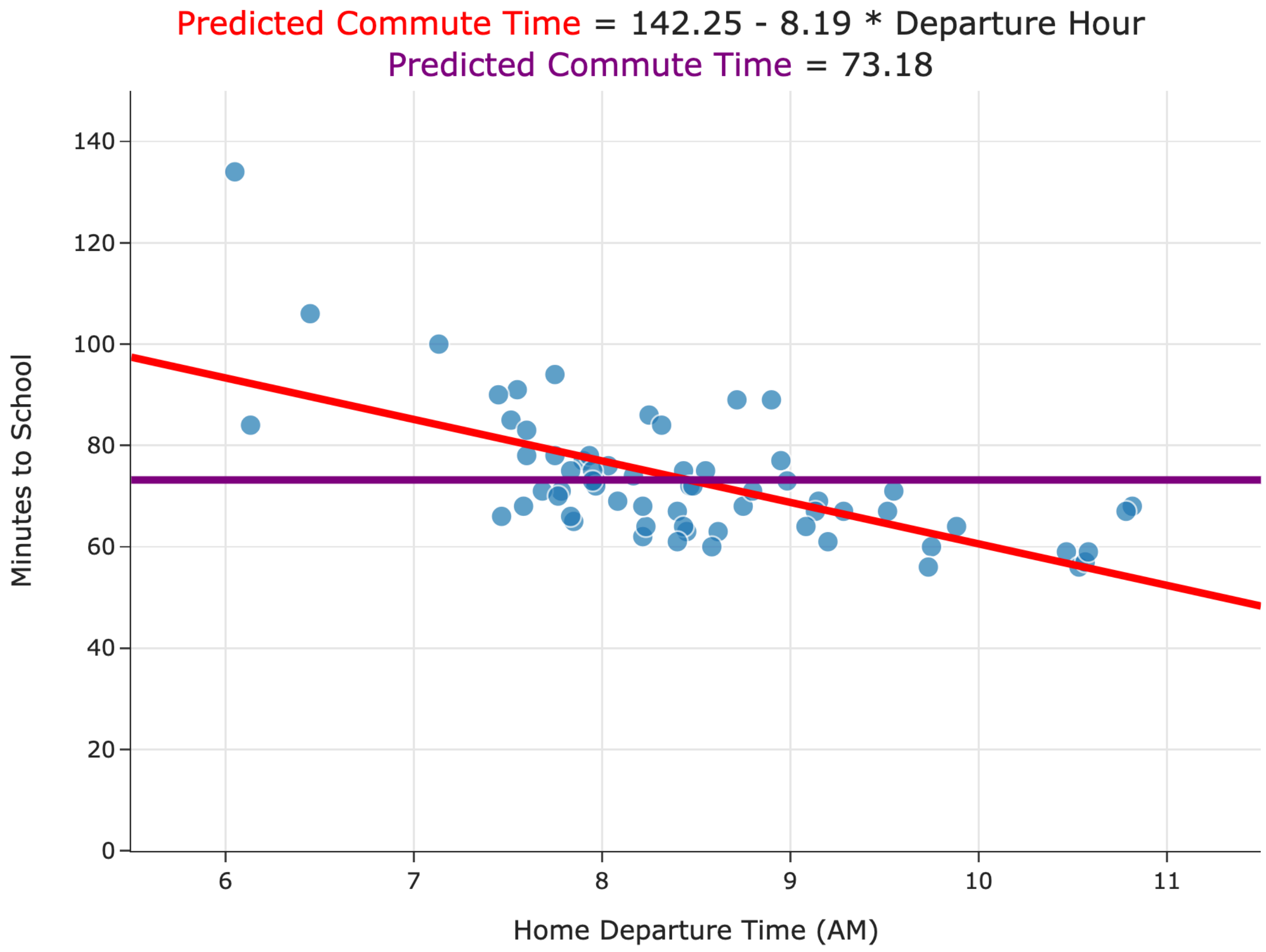
$$w_1^* = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

- If our model is $H(x) = w_0$, it is a **line that is forced to have a slope of** 0**, i.e. a horizontal line**. This is the same as the constant model from before.

$$w_0^* = \text{Mean}(y_1, y_2, \ldots, y_n)$$

- **Key idea**: w_0^* above is **not** necessarily equal to w_0^* for the simple linear regression model!

Comparing mean squared errors



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

- The MSE of the best simple linear regression model is ≈ 97.
- The MSE of the best constant model is ≈ 167.
- The simple linear regression model is a more flexible version of the constant model.

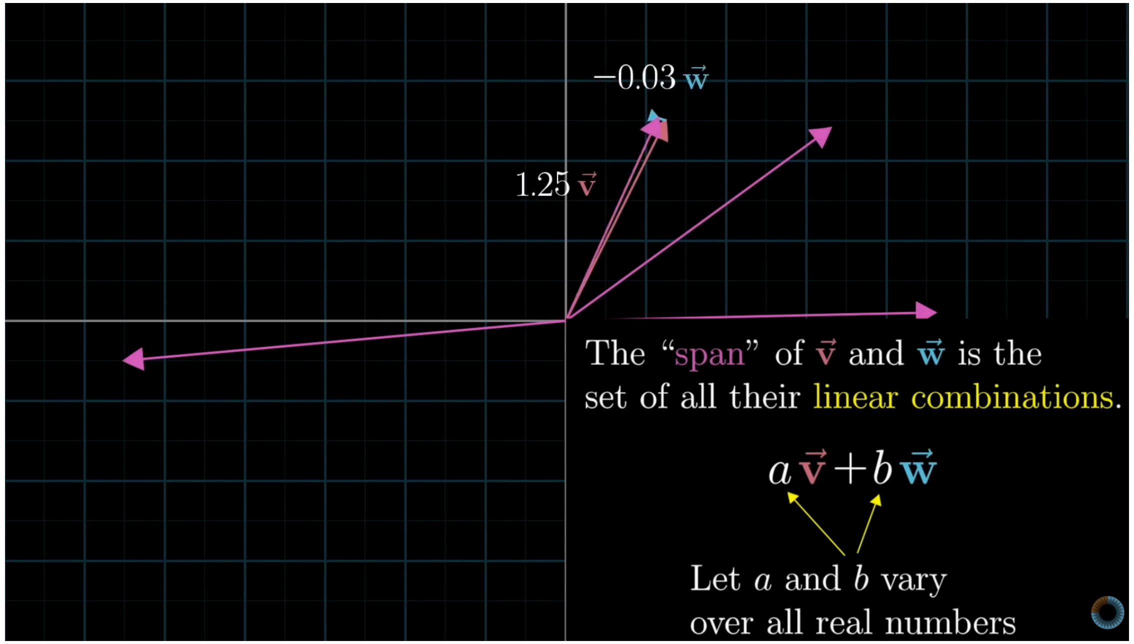
Dot products

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).
  - Are non-linear, e.g. $H(x) = w_0 + w_1 x + w_2 x^2$.
- 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.

Vectors

- 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**.

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

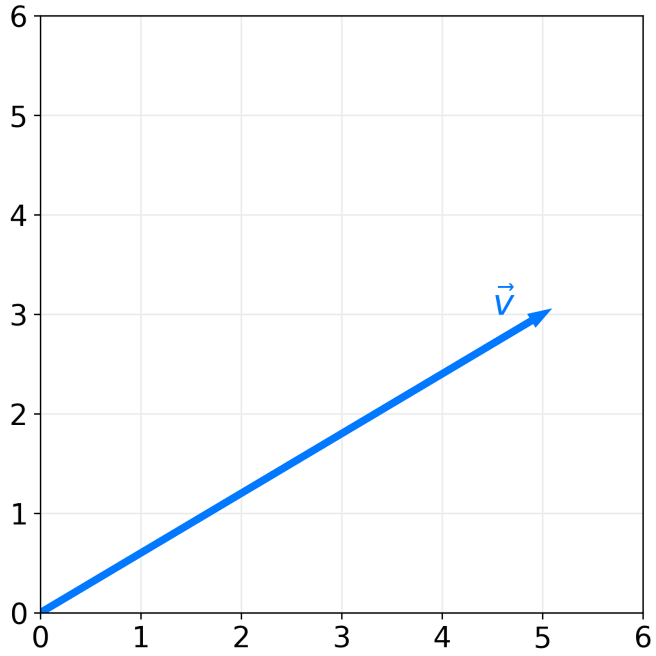
- Another way of writing the above vector is $\vec{v} = [8, 3, -2, 5]^T$.
- Since $\vec{v}$ has four **components**, we say $\vec{v} \in \mathbb{R}^4$.

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:

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

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



Dot product: coordinate definition

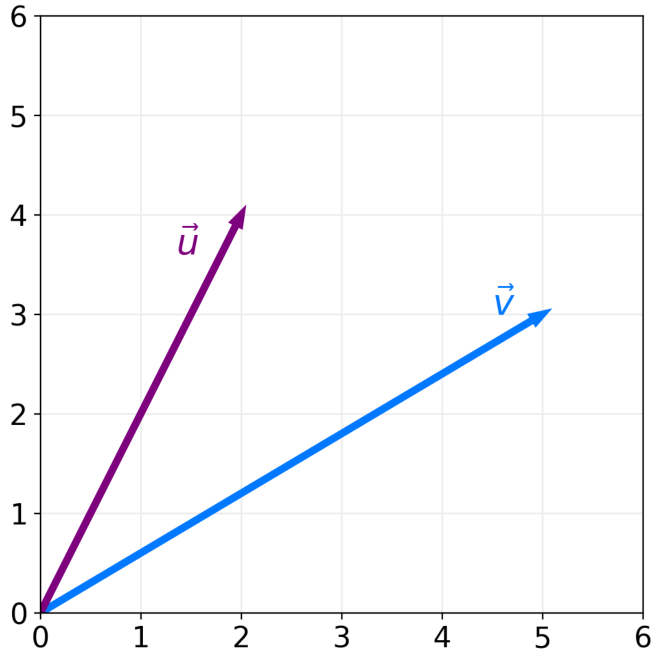
- The **dot product** of two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$ is written as:

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

- 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.



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.

Dot product: geometric definition

- The computational definition of the dot product:

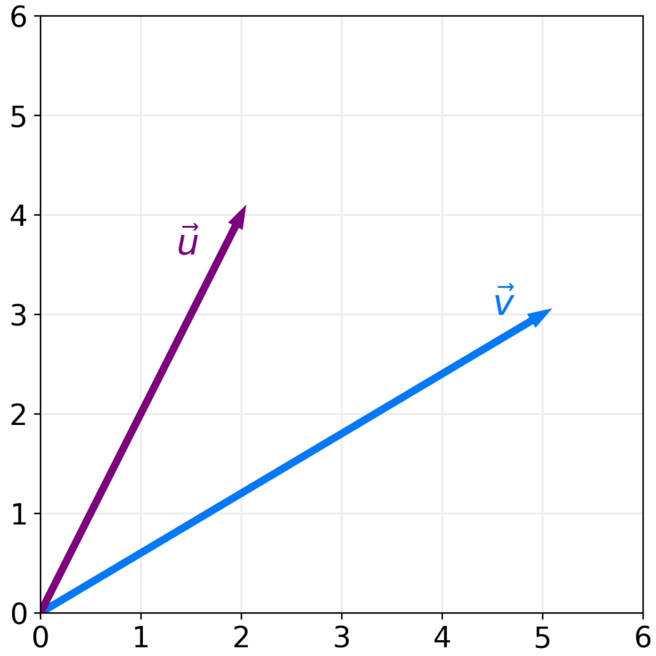
$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

- 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 $\vec{v}$.

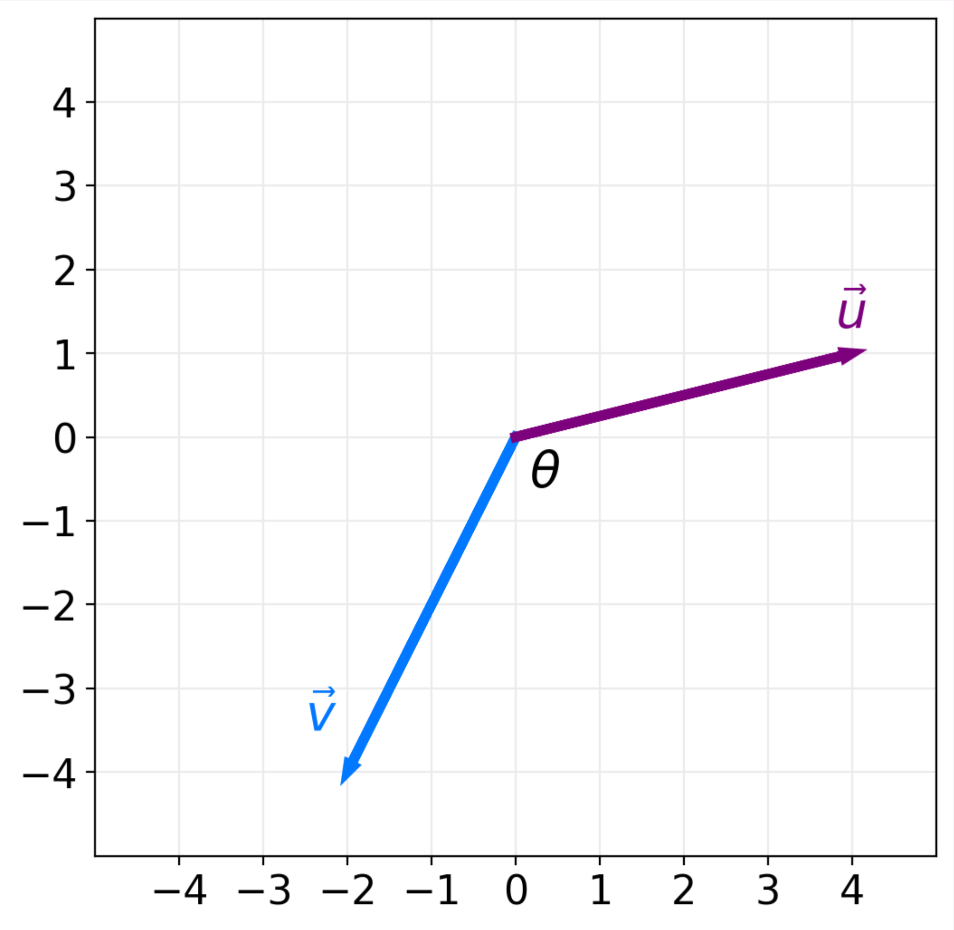
- The two definitions are equivalent! This equivalence allows us to find the angle θ between two vectors.



Question 🤔

Answer at q.dsc40a.com

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



Orthogonal vectors

- Recall: $\cos 90^{\circ} = 0$.
- Since $\vec{u} \cdot \vec{v} = ||\vec{u}|| \, ||\vec{v}|| \cos \theta$, 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**:

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

Exercise

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

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

Spans and projections

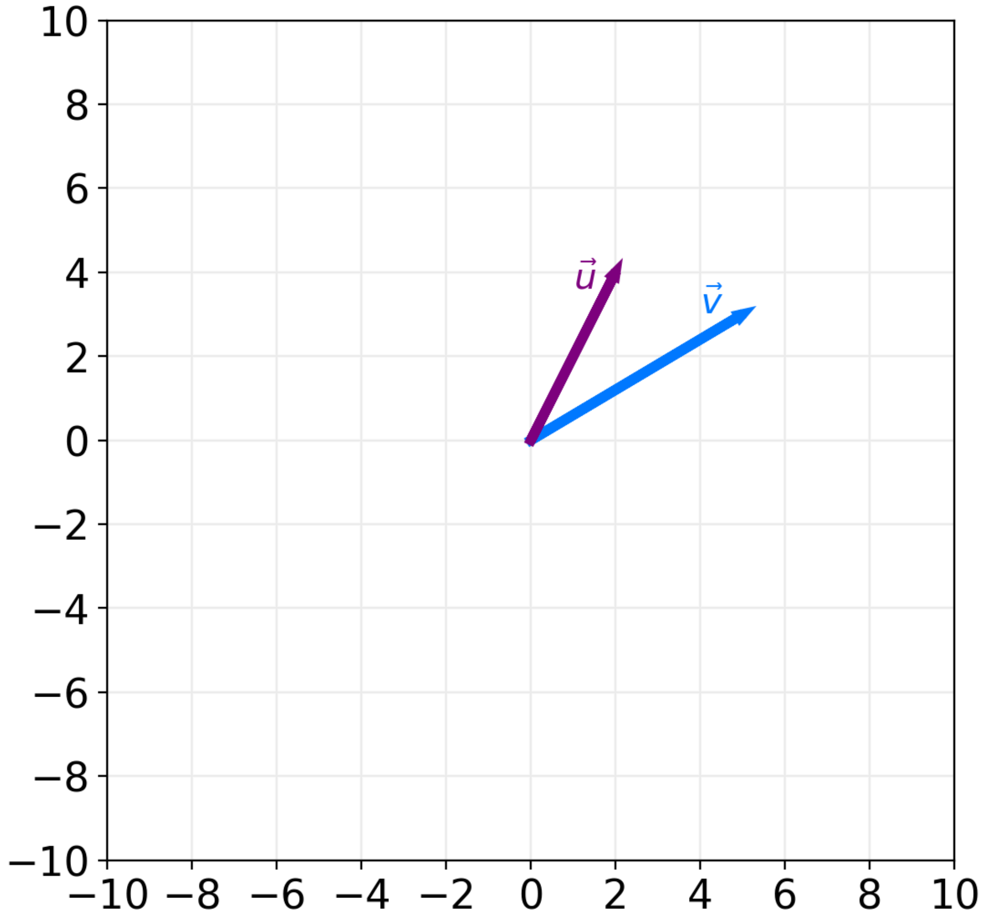
Adding and scaling vectors

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

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

- If c is a scalar, then:

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



Linear combinations

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

$$a_1\vec{v}_1 + a_2\vec{v}_2 + ... + a_n\vec{v}_n$$

where a_1, a_2, ..., a_n are all scalars.

Span

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

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

Exercise

Let $\vec{v}_1 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$ and let $\vec{v}_2 = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$. Is $\vec{y} = \begin{bmatrix} 9 \\ 1 \end{bmatrix}$ in $\mathrm{span}(\vec{v}_1, \vec{v}_2)$?

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

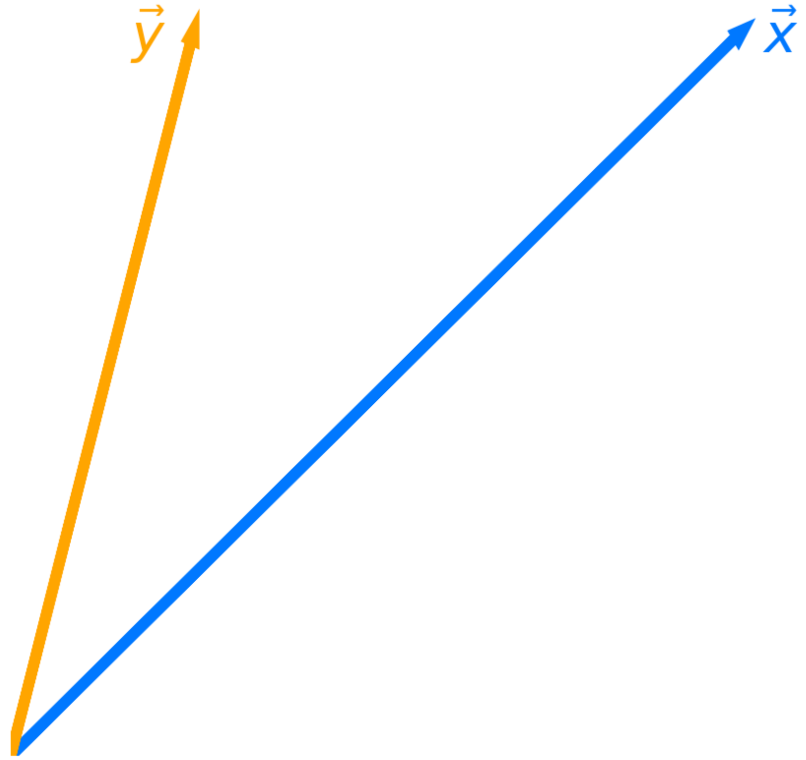
Projecting onto a single vector

- 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** $\text{span}(\vec{x})$.



Projection error

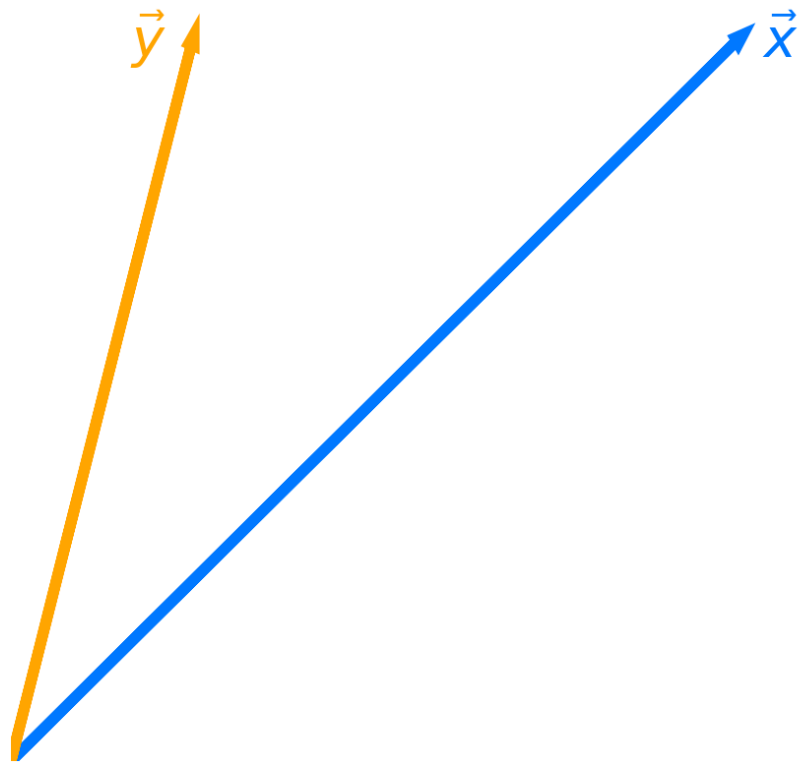
- Let $\vec{e} = \vec{y} - w\vec{x}$ be the **projection error**: that is, the vector that connects $\vec{y}$ to $\text{span}(\vec{x})$.
- **Goal: Find the w that makes $\vec{e}$ as short as possible.**
  - That is, minimize:

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

  - 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 $||\vec{e}||$?

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?

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:

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

- 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

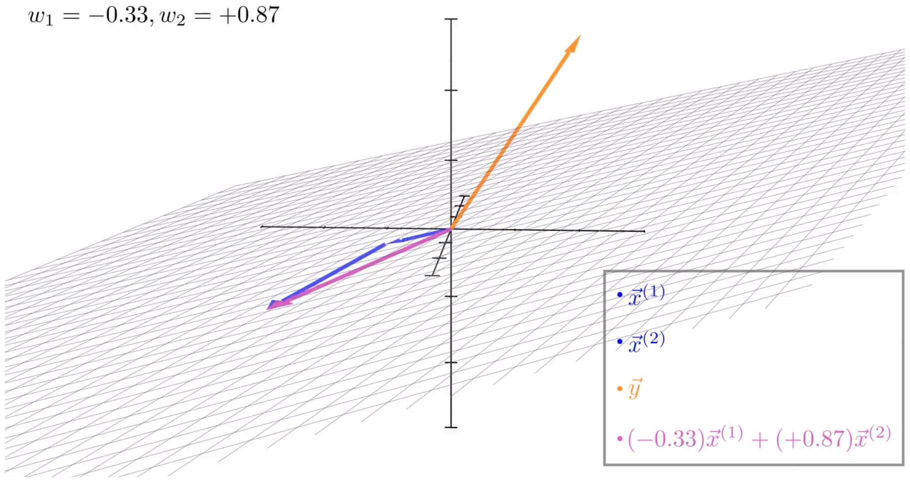
Let $\vec{a} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 9 \end{bmatrix}$.

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**, Xw.

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