Lecture 7

Orthogonal Projections

DSC 40A - UCSD

Agenda

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

Question 🤔

Take a moment to pause and reflect...

If you have any questions please post online to our forms/Q&A site.

Course staff will answer them ASAP!

Recap: Spans and projections

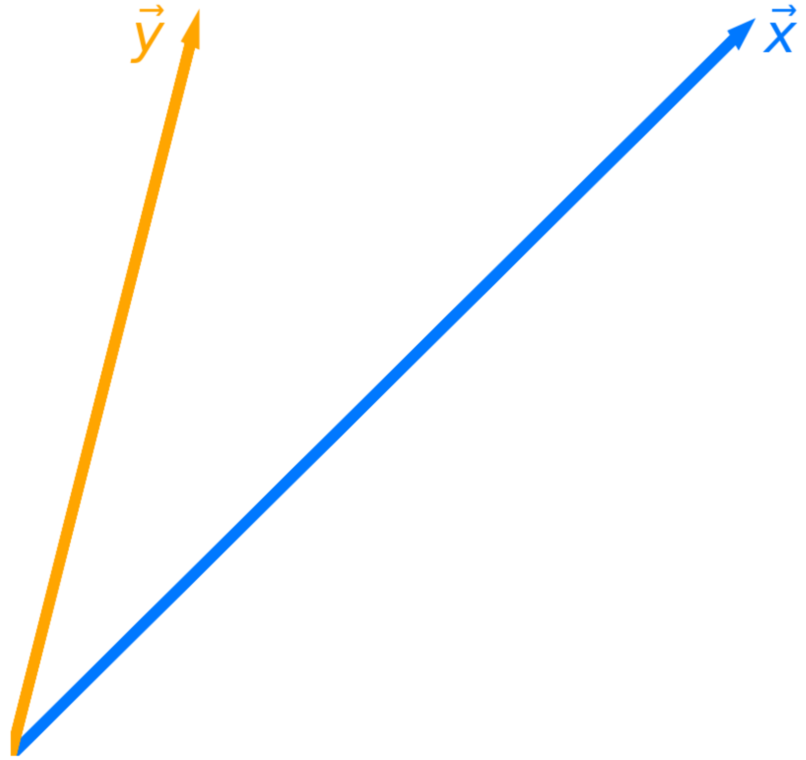
Projecting onto a single vector

- 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:

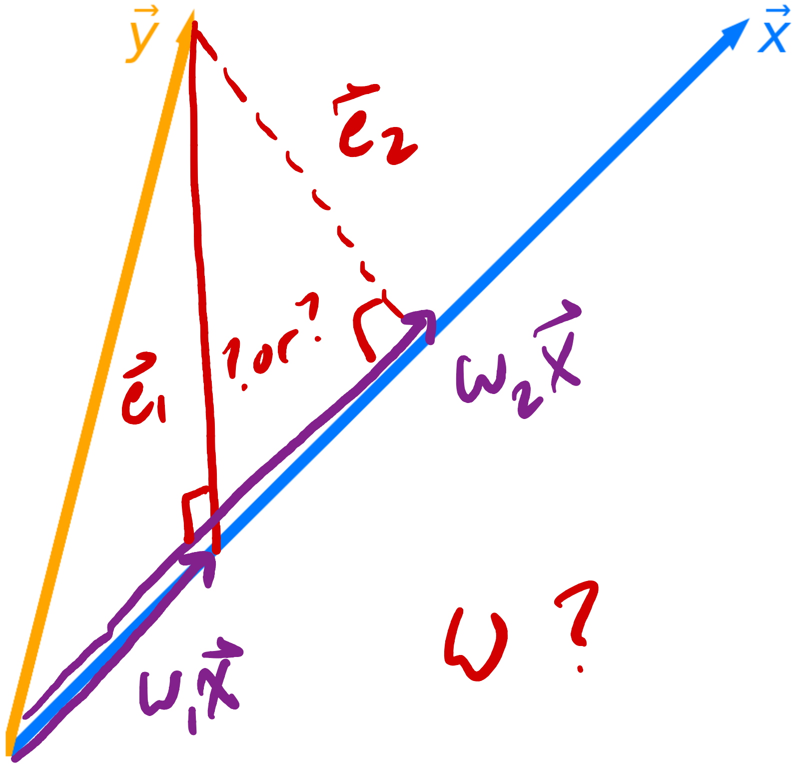
$$||\vec{e}||$$

  - Equivalently, minimize:

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

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

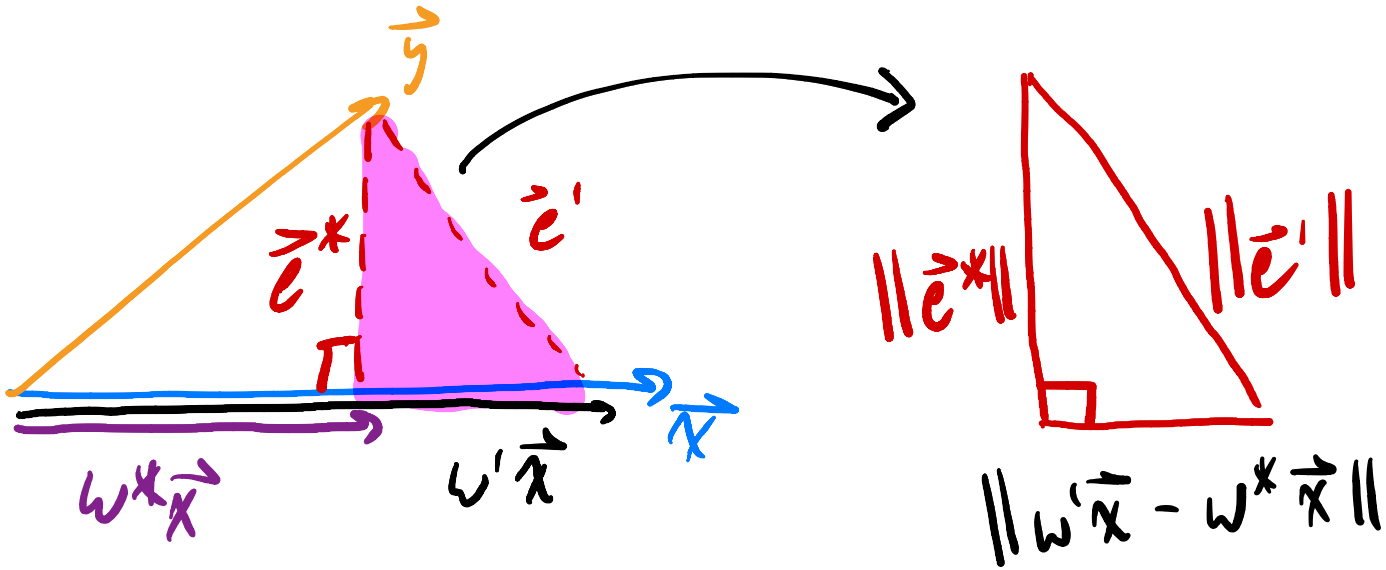
$\vec{e}$ is orthogonal to $\vec{x}$



Minimizing projection error

Purpose! show that $\vec{e}^*$ is the shortest possible $\vec{e}$ we can find

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



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

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

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

Minimizing projection error

$\vec{e}$ is orthogonal to $w\vec{x}$ $\left(w\vec{x} \cdot \vec{e} = 0\right)$

- **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?

w is a number

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

expand!

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

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

Distributive Property

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

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

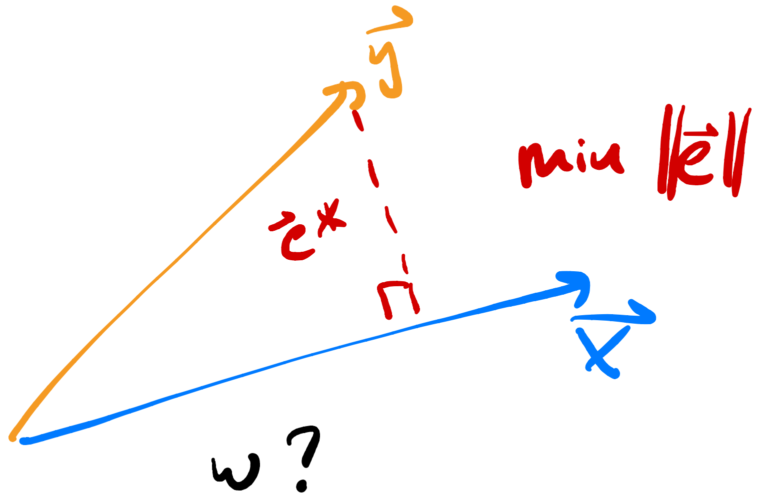
finds best scalar that projects $\vec{y}$ onto $\vec{x}$ by minimizing the magnitude of $\vec{e}$

i.e. w^* gives us our best or shortest $\vec{e}$

Orthogonal projection

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

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



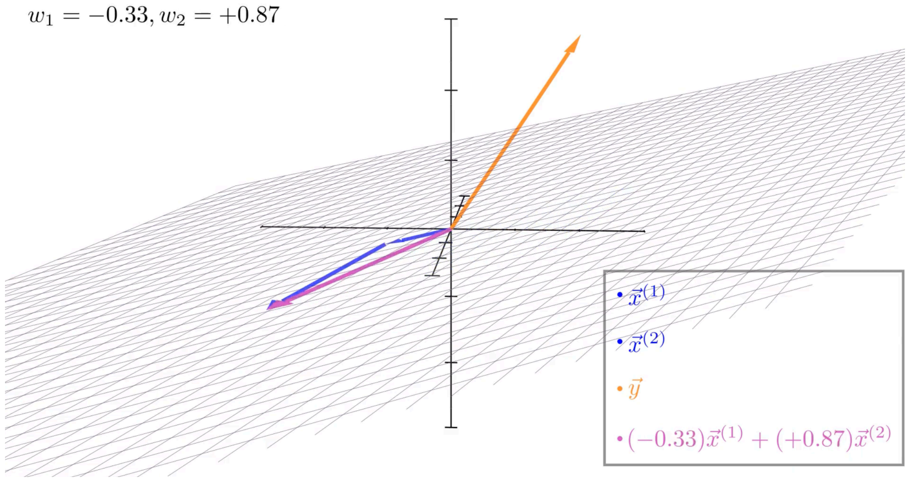
- Note that w^* is the solution to a minimization problem, specifically, this one:

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

- 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 $\vec{y}$.

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:

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

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:

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

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 $\vec{e}$:**

Trying to find

w_1^* , w_2^*

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

- 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.

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

- 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**.

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

Matrix addition and scalar multiplication

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

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

$2 \times 3 \qquad 2 \times 3$

- Scalar multiplication occurs elementwise, too:

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

$\boxed{A = 2\times3}$ 2×3 2×3

? X can't Do it

Matrix-matrix multiplication

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

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

- 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^TA?

$$A = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 5 & -3 \end{bmatrix} \quad 2\times3$$

$A \quad B$

$n\times d \quad d\times p$

need to match!

$$A^T = \begin{bmatrix} 2 & -1 \\ 5 & 5 \\ 8 & -3 \end{bmatrix} \quad 3\times2$$

$$A^TA = \begin{bmatrix} 5 & \_ & \_ \\ 5 & \_ & \_ \\ 19 & \_ & \boxed{73} \end{bmatrix} \quad 3\times3$$

$[2\ -1]\cdot\begin{bmatrix}2\\-1\end{bmatrix} = 4+1=5$

$[5\ 5]\begin{bmatrix}2\\-1\end{bmatrix} = 5$

$[8\ -3]\cdot\begin{bmatrix}2\\-1\end{bmatrix} = 19$

$[8\ -3]\cdot\begin{bmatrix}8\\-3\end{bmatrix} = 73$

Question 🤔

Take a moment to pause and reflect...

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

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

$$A_{5\times 7}\, B_{7\times 5} \rightarrow 5\times 5$$

$$B_{7\times 5}\, A_{5\times 7} \rightarrow 7\times 7$$

Diff. Dimensions

Matrix-vector multiplication

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

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

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

$$A_{n \times d} \vec{v}_{d \times 1} = \vec{v} \in \mathbb{R}^d \qquad d \text{ components}$$

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

$$\vec{v}^T_{1 \times n} A_{n \times d} = \vec{v} \in \mathbb{R}^n \qquad n \text{ components}$$

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 $A\vec{v}$?

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

$\vec{v} \in \mathbb{R}^3$

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

$A\vec{v} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}_{2\times 1}$

$\underset{2\times 3}{A} \quad \underset{3\times 1}{\vec{v}}$

I. $2(2) + (-1)(5) + (-5)(8)$

$4 \quad -5 \quad -40 \quad = -41$

II. $2(-1) + (-1)(5) + (-5)(-3)$

$-2 + -5 + 15 \quad = 8$

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 $\vec{v}$.**

$w_1\vec{x}^1 + w_2\vec{x}^2 + w_3\vec{x}^3$

- Example: What is $A\vec{v}$?

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

$$A\vec{v} = \underbrace{(2)\begin{bmatrix} 2 \\ -1 \end{bmatrix} + (-1)\begin{bmatrix} 5 \\ 5 \end{bmatrix} + (-5)\begin{bmatrix} 8 \\ -3 \end{bmatrix}}_{\text{Linear Combination}} = \begin{bmatrix} -41 \\ 8 \end{bmatrix}$$

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

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

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

$$\vec{x}^{(1)} \cdot \vec{e} = 0$$

$$\vec{x}^{(2)} \cdot \vec{e} = 0$$

Matrix-vector products create linear combinations of columns!

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

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

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

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

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

Equal

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

- 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:

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

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

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

- **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

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

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

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

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

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

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

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

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

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

$$X^T = \begin{bmatrix} 2 & 5 & 3 \\ -1 & 0 & 4 \end{bmatrix}_{2\times 3} \qquad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{3\times 1}$$

$$\Rightarrow X^T\vec{e} = \begin{bmatrix} \text{—} \; \vec{x}^{(1)T} \; \text{—} \\ \text{—} \; \vec{x}^{(2)T} \; \text{—} \end{bmatrix} \vec{e}$$

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

The normal equations

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

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

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

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

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

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

How? Last few slides.

normal = orthogonal

is one equation a system of linear equations

- 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:

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

whole Point of lecture was to determine but How we come to calculate these values

- 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!