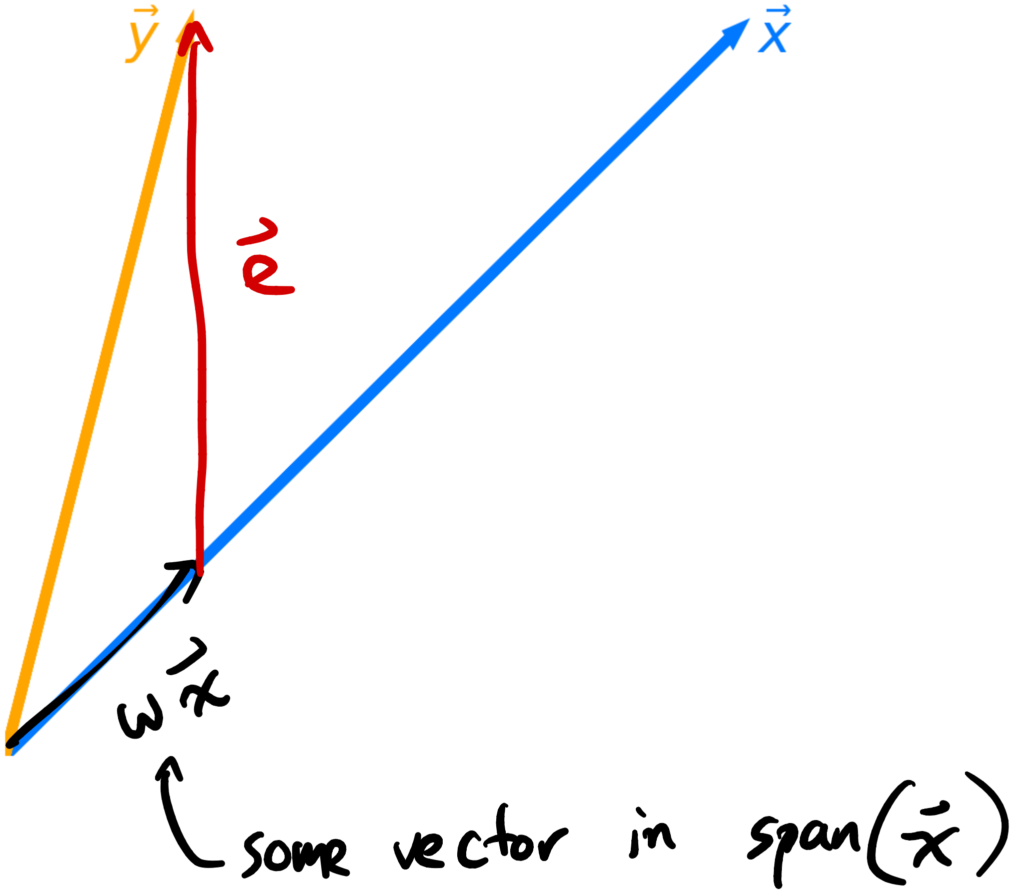
Overview: Spans and projections

Projecting onto the span of a single vector

- **Question**: What vector in $\text{span}(\vec{x})$ is closest to $\vec{y}$?
- The answer is the vector $w\vec{x}$, where the w is chosen to minimize the **length** of the error vector:

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

- **Key idea**: To minimize the length of the error vector, choose w so that the error vector is **orthogonal** to $\vec{x}$.



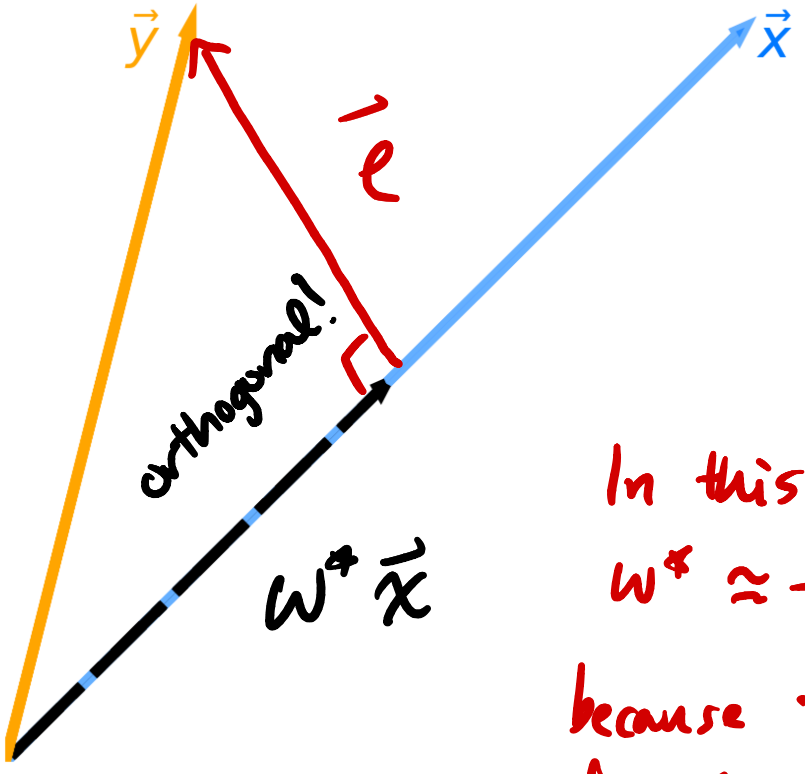
Projecting onto the span of a single vector

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

$$w^* = \underbrace{\frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}}}_{\text{a scalar!}}$$

How did we find w^*?

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



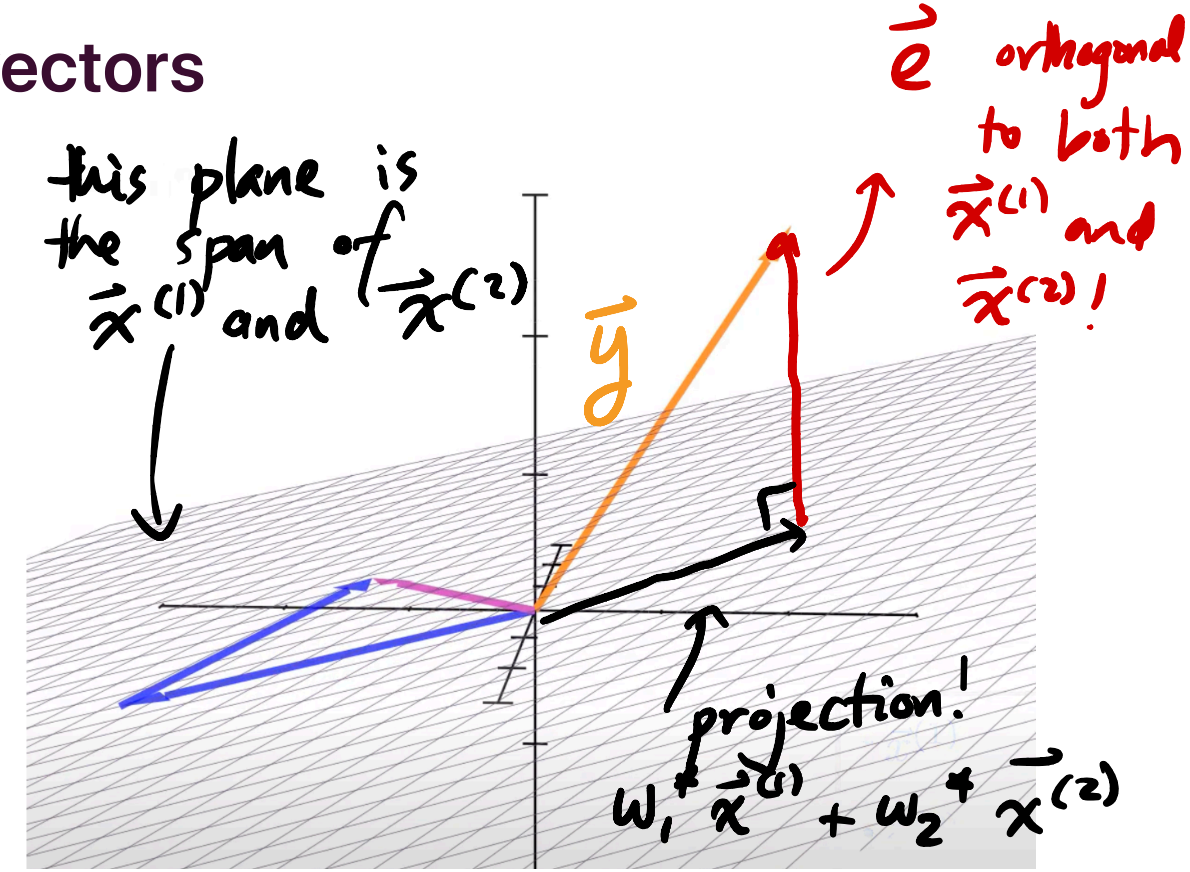
In this example, $w^* \approx \frac{1}{2}$, because the length of $w^* \vec{x}$ is $\approx$ $\frac{1}{2}$ the length of $\vec{x}$.

Projecting onto the span of multiple vectors

- **Question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
- The answer is the vector $w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}$, where w_1 and w_2 are chosen to minimize the **length** of the error vector:

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

- **Key idea**: To minimize the length of the error vector, choose w_1 and w_2 so that the error vector is **orthogonal** to **both** $\vec{x}^{(1)}$ **and** $\vec{x}^{(2)}$.



If $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ are **linearly independent**, they span a **plane**.

Matrix-vector products create linear combinations of columns!

- **Question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
- To help, we can create a **matrix**, X, by stacking $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ next to each other:

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

- Then, instead of writing vectors in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ as $w_1 \vec{x}^{(1)} + w_2 \vec{x}^{(2)}$, we can say:

$$X\vec{w} \qquad \text{where } \vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

- **Key idea**: Find $\vec{w}$ such that the error vector, $\vec{e} = \vec{y} - \underbrace{X\vec{w}}$, is **orthogonal** to **every column of** X.

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

$A\vec{v}$ ← the dot product of $\vec{v}$ with every <u>row</u> of A

Constructing an orthogonal error vector

- **Key idea**: Find $\vec{w} \in \mathbb{R}^d$ such that the error vector, $\vec{e} = \vec{y} - X\vec{w}$, is **orthogonal** to the **columns of** X.
  - Why? Because this will make the error vector as short as possible.
- The $\vec{w}^*$ that accomplishes this satisfies:

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

$$\underbrace{X^T\vec{e}}_{\text{a vector!}} = \vec{0} \leftarrow \text{the zero vector!}$$

  - Why? Because $X^T\vec{e}$ contains the **dot products** of each column in X with $\vec{e}$. If these are all 0, then $\vec{e}$ is **orthogonal** to **every column of** X!

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

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

↙ this is just $\vec{x}^{(1)} \cdot e$!

Aside
$$\left(\tfrac{1}{2}\right)2x = \left(\tfrac{1}{2}\right)5$$
$$x = 5/2$$

The normal equations

- **Key idea**: Find $\vec{w} \in \mathbb{R}^d$ such that the error vector, $\vec{e} = \vec{y} - X\vec{w}$, is **orthogonal** to the **columns of** X.

- The $\vec{w}^*$ that accomplishes this satisfies:

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

system of equations

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

- 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**.

all columns are linearly independent

  - If X^TX is not full rank, then there are infinitely many solutions to the normal equations, $X^TX\vec{w}^* = X^T\vec{y}$.

equivalent: X is full rank

What does it mean?

- **Original question**: What vector in $\text{span}(\vec{x}^{(1)}, \vec{x}^{(2)})$ is closest to $\vec{y}$?
- **Final answer**: Assuming X^TX is invertible, 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 one that satisfies the **normal equations**:

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

  If $X^T X$ is invertible, then the unique solution is:

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

- We're going to use this frequently!