Spans and projections, revisited

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

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 and w_2 such that:

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

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

Matrix-vector products create linear combinations of columns!

the same!

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

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

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

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

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

the 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

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

- **Goal**: Find the vector $\vec{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^T$ such that $\|\vec{e}\| = \|\vec{y} - \underbrace{X\vec{w}}_{w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)}}\|$ is minimized.
- As we've seen, $\vec{w}$ must be such that:

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

- **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

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

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

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

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

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

$$\Rightarrow \boxed{\begin{aligned} \vec{x}^{(1)} \cdot (\vec{y} - X\vec{w}) &= 0 \\ \vec{x}^{(2)} \cdot (\vec{y} - X\vec{w}) &= 0 \end{aligned}}$$

Simplifying the system of equations, using matrices

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

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

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

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

$\Downarrow$ combine into a single equation

$$X^T \underbrace{(\vec{y} - X\vec{w})}_{\vec{e}} = \vec{0}$$

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

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

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

example

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

$\vec{x} \cdot \vec{y}$
$= \vec{x}^T \vec{y}$

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

rows of X^T are the columns of X!!!

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 $\vec{w} = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^T$ such that $||\vec{e}|| = ||\vec{y} - X\vec{w}||$ is minimized.
- We now know that it is the vector $\vec{w}^*$ such that:

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

previous slide

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

The general solution to the normal equation

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

- **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:

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

- 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 $0.7289\vec{x}^{(1)} + 1.6300\vec{x}^{(2)}$.

Overview: Spans and projections

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**:

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

  If X^TX is invertible, then the unique solution is:

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

- We're going to use this frequently!