Math Refresher for DS

Practical Session 4

Plan for Today

- A short quiz
- SLE with no solutions
- Practice in Python

We (Finally) Have a Course Repo!

- https://github.com/girafe-ai/math-basics-for-ai
    - Slides
    - Links to colab-notebooks
    - Links to lectures / practical session recordings
    - Additional material

Short Quiz Lectures 1 - 3

https://forms.gle/Mw28SUSTwohWWv9d7

Solving Systems of Linear Equations

- $Ax = b$ – a system of linear equations (SLE).
- $A - m \times n$ matrix (= m equations, n variables).

Solving Systems of Linear Equations

- $Ax = b$ – a system of linear equations (SLE).
- $A - m \times n$ matrix (= m equations, n variables).

- How can we determine the number of solutions to it?

Solving Systems of Linear Equations

- $Ax = b$ – a system of linear equations (SLE).
- $A - m \times n$ matrix (= m equations, n variables).

- How can we determine the number of solutions to it?
  - $rank(A) = rank(A|b) = n \Leftrightarrow$ there is a unique solution $x = A^{-1}b$;
  - $rank(A) = rank(A|b) < n \Leftrightarrow$ there are infinitely many solutions;
  - $rank(A) < rank(A|b) \Leftrightarrow$ there are no solutions.

girafe ai

Solving Systems of Linear Equations

- $Ax = b$ – a system of linear equations (SLE).
- $A - m \times n$ matrix (= m equations, n variables).

- How can we determine the number of solutions to it?
  - $rank(A) = rank(A|b) = n \Leftrightarrow$ there is a unique solution $x = A^{-1}b$;
  - $rank(A) = rank(A|b) < n \Leftrightarrow$ there are infinitely many solutions;
  - $rank(A) < rank(A|b) \Leftrightarrow$ there are no solutions.

- Sometimes, we need to "solve" systems that don't have a solution.

Solving Systems of Linear Equations

- $Ax = b$ – a system of linear equations (SLE).
- $A - m \times n$ matrix (= m equations, n variables).

- How can we determine the number of solutions to it?
  - $rank(A) = rank(A|b) = n \Leftrightarrow$ there is a unique solution $x = A^{-1}b$;
  - $rank(A) = rank(A|b) < n \Leftrightarrow$ there are infinitely many solutions;
  - $rank(A) < rank(A|b) \Leftrightarrow$ there are no solutions.

- Sometimes, we need to "solve" systems that don't have a solution.

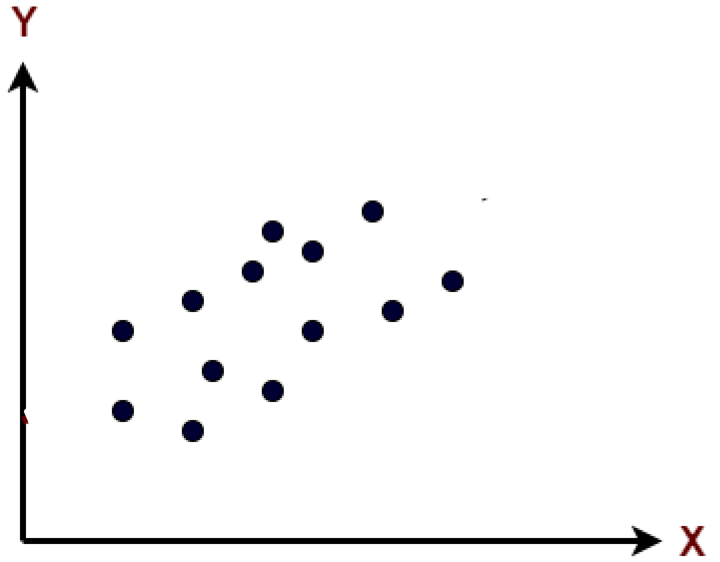
How to find a reasonable approximate solution?

Least Squares

Motivating Example

- Imagine that you have m observations:

$$(x_1, y_1), \qquad (x_2, y_2), \qquad \dots, \qquad (x_m, y_m)$$

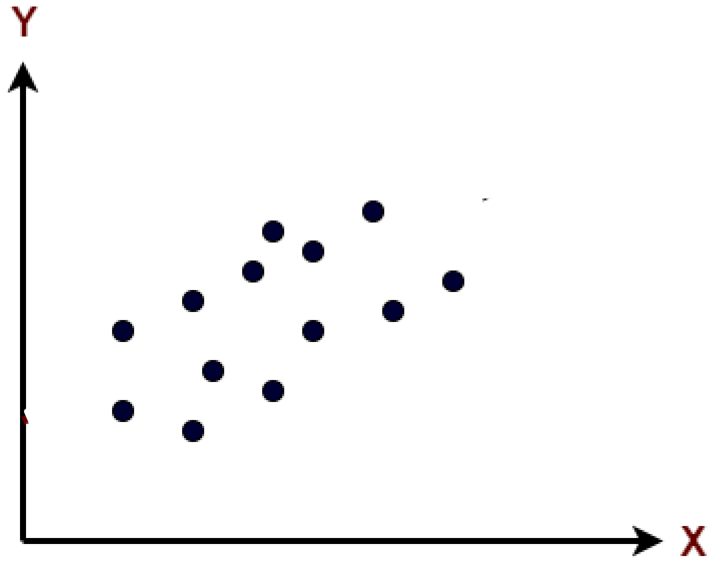


Motivating Example

- Imagine that you have m observations:

$$(x_1, y_1), \quad (x_2, y_2), \quad \dots, \quad (x_m, y_m)$$

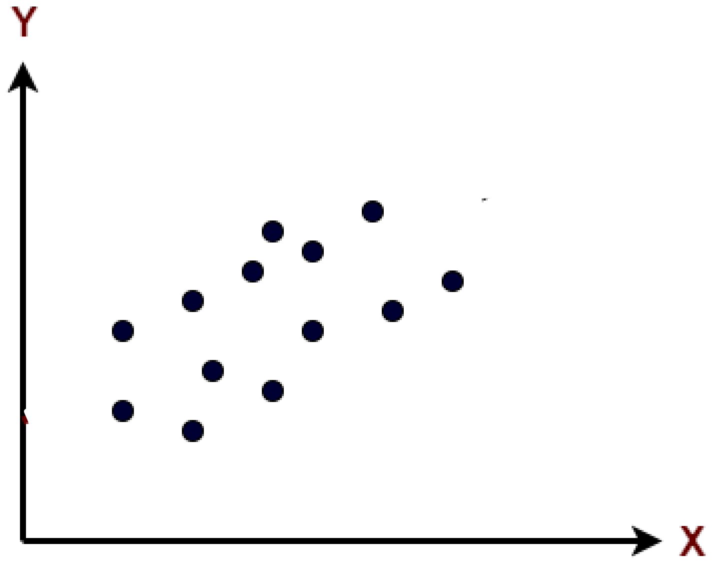
- You want to draw a line through your observations:



Motivating Example

- Imagine that you have m observations:

$$(x_1, y_1), \quad (x_2, y_2), \quad \dots, \quad (x_m, y_m)$$



- You want to draw a line through your observations:

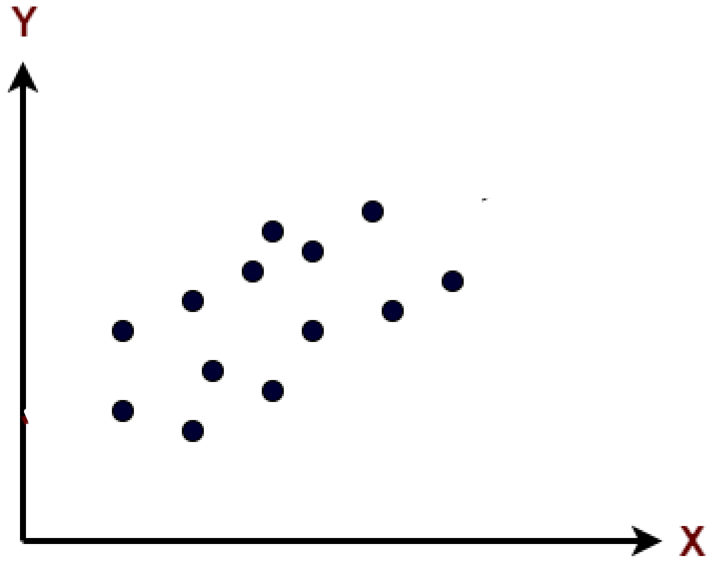
$$\begin{matrix} \boldsymbol{w_0} + \boldsymbol{w_1} x_1 = y_1 \\ \boldsymbol{w_0} + \boldsymbol{w_1} x_2 = y_n \\ \vdots \\ \boldsymbol{w_0} + \boldsymbol{w_1} x_m = y_m \end{matrix} \quad \Leftrightarrow$$

Motivating Example

- Imagine that you have m observations:

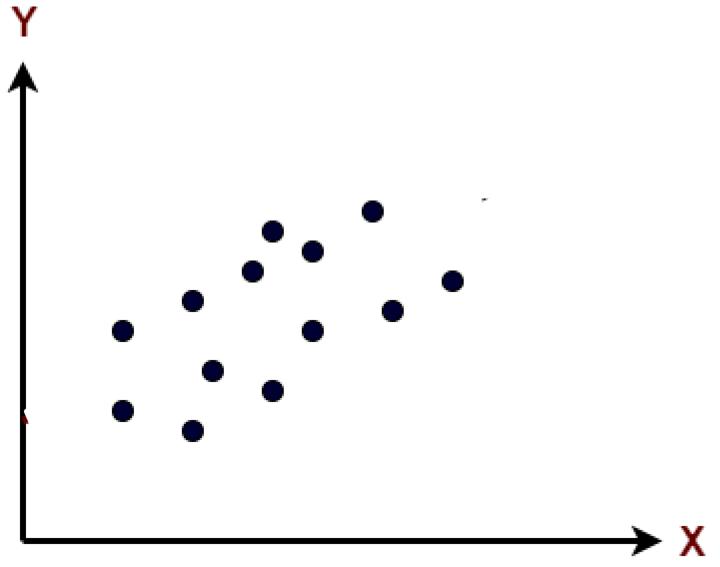
$$(x_1, y_1), \quad (x_2, y_2), \quad \dots, \quad (x_m, y_m)$$

- You want to draw a line through your observations:



$$\begin{gathered} w_0 + w_1 x_1 = y_1 \\ w_0 + w_1 x_2 = y_n \\ \vdots \\ w_0 + w_1 x_m = y_m \end{gathered} \iff Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Motivating Example



- Imagine that you have m observations:

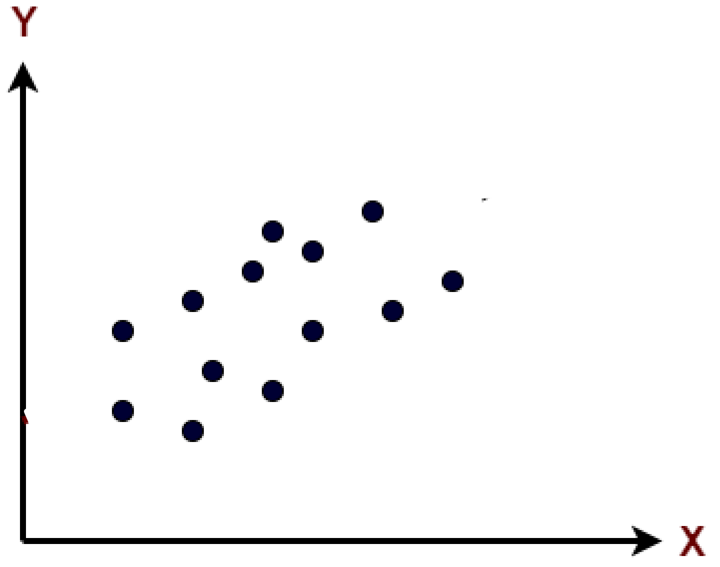
$$(x_1, y_1), \qquad (x_2, y_2), \qquad \dots, \qquad (x_m, y_m)$$

- You want to draw a line through your observations:

$$\begin{matrix} w_0 + w_1 x_1 = y_1 \\ w_0 + w_1 x_2 = y_n \\ \vdots \\ w_0 + w_1 x_m = y_m \end{matrix} \iff Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

- Overdetermined system, no solutions.

Motivating Example



- Imagine that you have m observations:

$$(x_1, y_1), \qquad (x_2, y_2), \qquad \dots, \qquad (x_m, y_m)$$

- You want to draw a line through your observations:

$$\begin{array}{c} w_0 + w_1 x_1 = y_1 \\ w_0 + w_1 x_2 = y_n \\ \vdots \\ w_0 + w_1 x_m = y_m \end{array} \iff Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

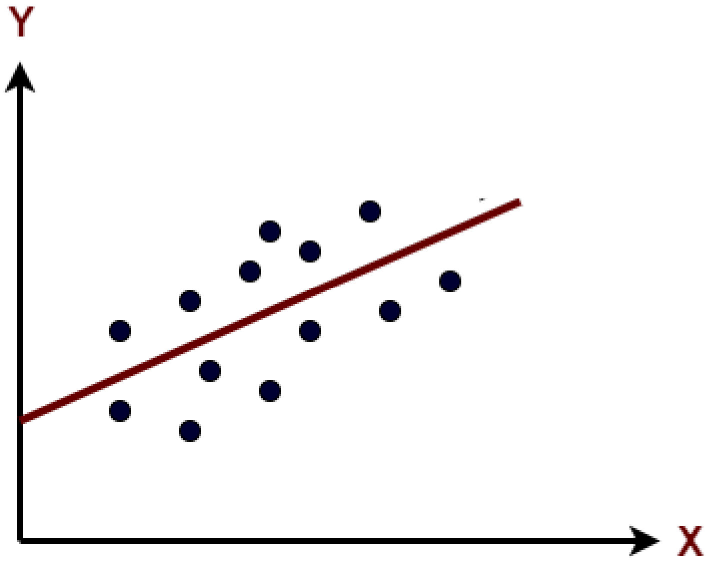
- Overdetermined system, no solutions.

- What shall we do?

Motivating Example

- Imagine that you have m observations:

$$(x_1, y_1), \quad (x_2, y_2), \quad \dots, \quad (x_m, y_m)$$



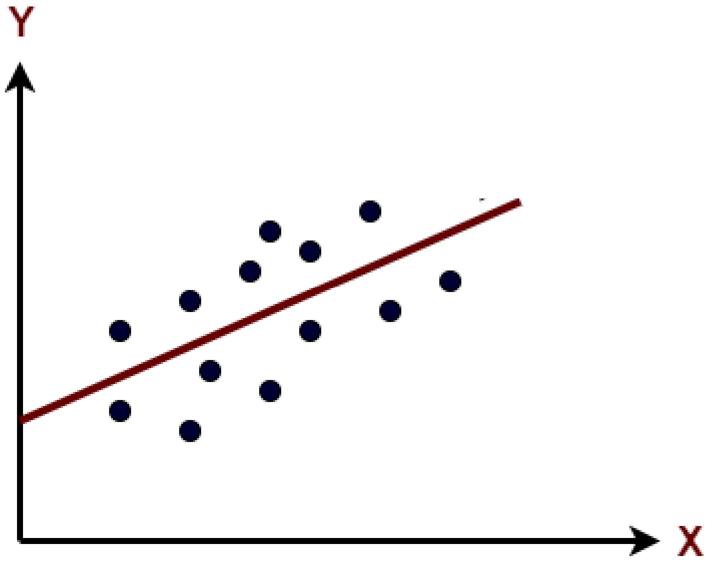
- You want to draw a line through your observations:

$$\begin{matrix} w_0 + w_1 x_1 = y_1 \\ w_0 + w_1 x_2 = y_n \\ \vdots \\ w_0 + w_1 x_m = y_m \end{matrix} \iff Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

- Overdetermined system, no solutions.

- What shall we do?

Motivating Example



- Imagine that you have m observations:

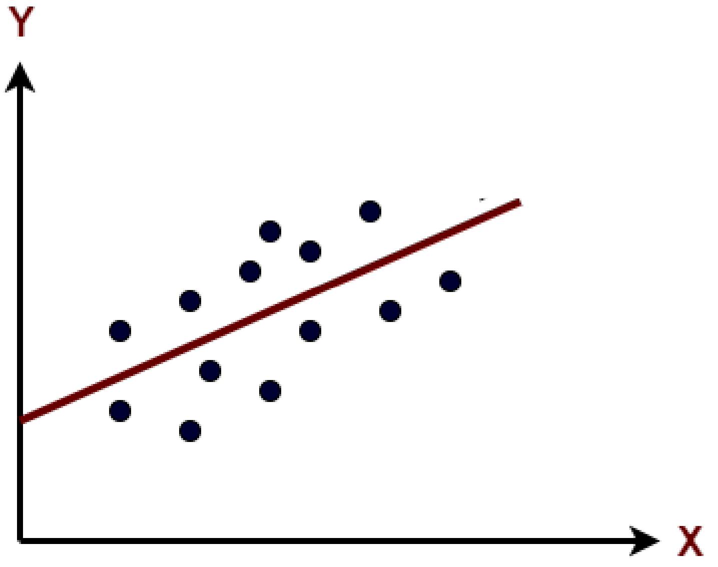
$$(x_1, y_1), \qquad (x_2, y_2), \qquad \dots, \qquad (x_m, y_m)$$

- You want to draw a line through your observations:

$$\begin{aligned} w_0 + w_1 x_1 &= y_1 \\ w_0 + w_1 x_2 &= y_n \\ &\vdots \\ w_0 + w_1 x_m &= y_m \end{aligned} \iff Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

- Overdetermined system, no solutions.
- What shall we do? Method of least squares: pick a line such that average squared error is minimized.

Motivating Example



- Imagine that you have m observations:

$$(x_1, y_1), \qquad (x_2, y_2), \qquad \dots, \qquad (x_m, y_m)$$

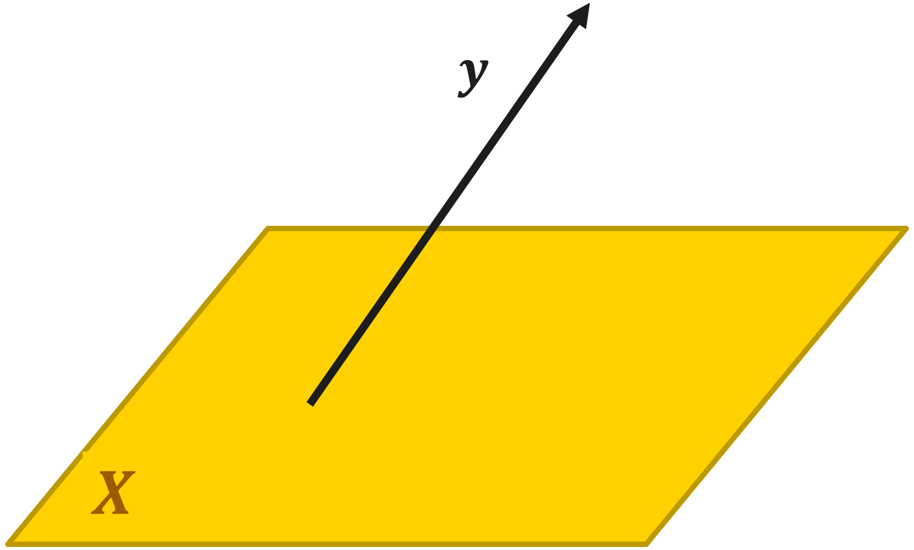
- You want to draw a line through your observations:

$$\begin{matrix} w_0 + w_1 x_1 = y_1 \\ w_0 + w_1 x_2 = y_n \\ \vdots \\ w_0 + w_1 x_m = y_m \end{matrix} \iff Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

- Overdetermined system, no solutions.
- What shall we do? Method of least squares: pick a line such that average squared error is minimized.
- Let's look at it from the Linear Algebra perspective.

Method of Least Squares

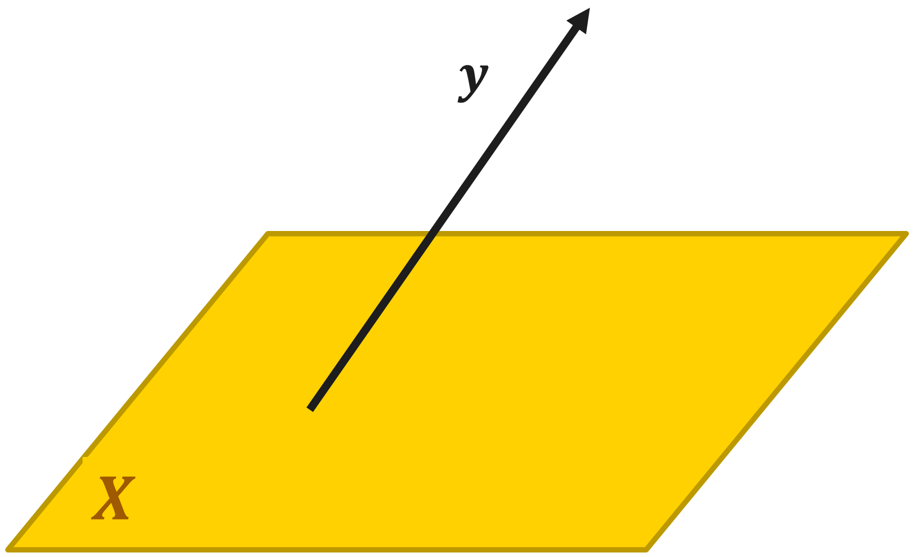
- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.



Method of Least Squares

- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.

- Let's chose $\hat{y}$ such that



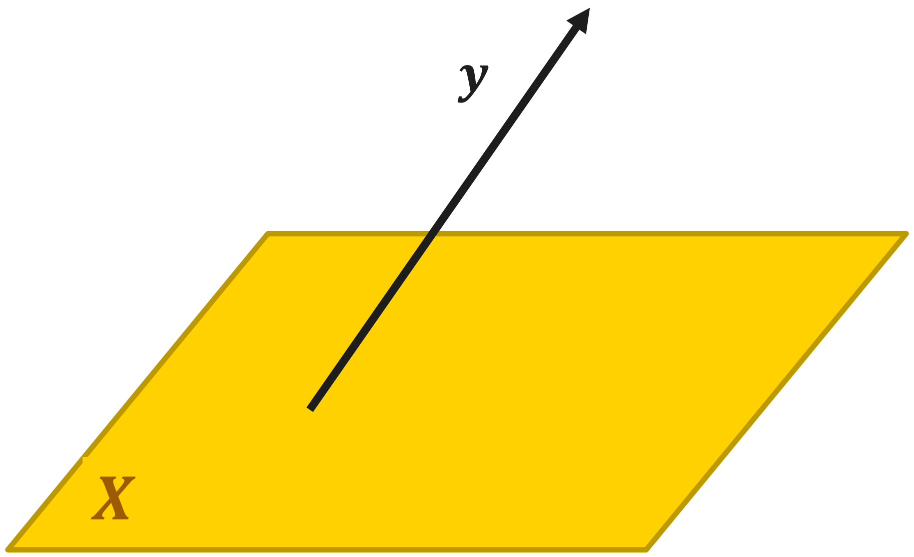
Method of Least Squares

- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.

- Let's chose $\hat{y}$ such that

  $Xw = \hat{y}$ has a solution w^*

  $\Leftrightarrow$



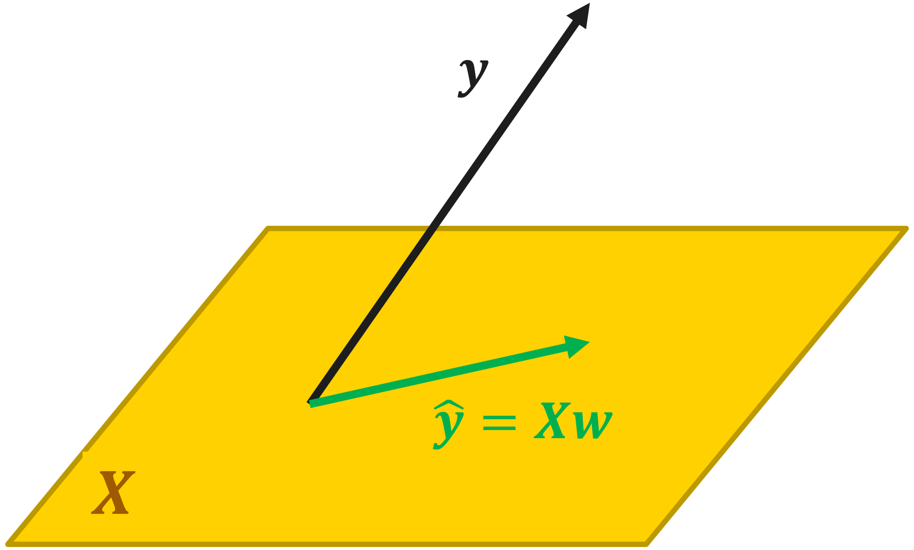
Method of Least Squares

- $Xw = y$ has no solutions $\Leftrightarrow$ y is **not** in the column space of X.

- Let's chose $\hat{y}$ such that

  $Xw = \hat{y}$ has a solution w^*
  $\Leftrightarrow$ $\hat{y}$ is in the column space of X

  and



Method of Least Squares

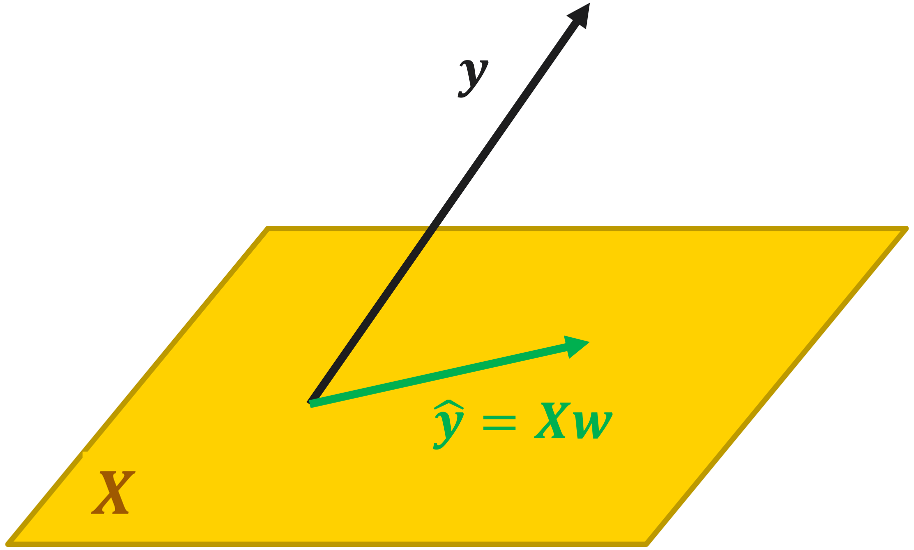
- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.

- Let's chose $\hat{y}$ such that

  $Xw = \hat{y}$ has a solution w^*
  $\Leftrightarrow$ $\hat{y}$ is in the column space of X

  and

  $\hat{y}$ is as close to y as possible
  $\Leftrightarrow$



Method of Least Squares

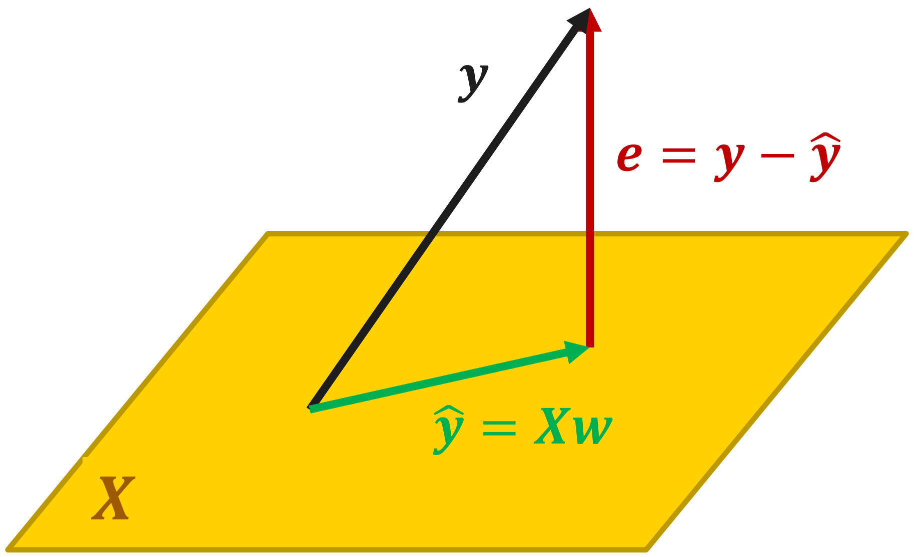
- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.

- Let's chose $\hat{y}$ such that

  $Xw = \hat{y}$ has a solution w^*
  $\Leftrightarrow$ $\hat{y}$ is in the column space of X

  and

  $\hat{y}$ is as close to y as possible
  $\Leftrightarrow$ $\|y - \hat{y}\|$ is minimized



Method of Least Squares

- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.

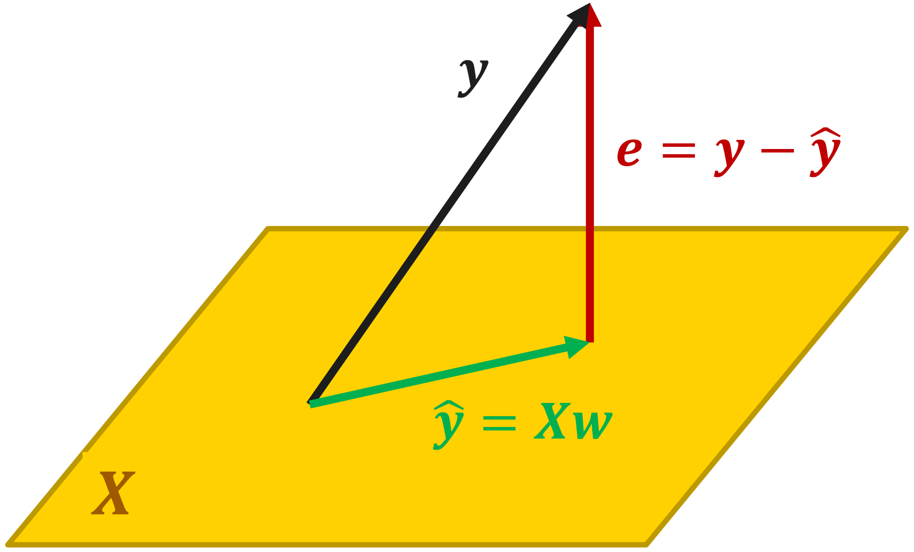
- Let's chose $\hat{y}$ such that

  $Xw = \hat{y}$ has a solution w^*
  $\Leftrightarrow$ $\hat{y}$ is in the column space of X

  and

  $\hat{y}$ is as close to y as possible
  $\Leftrightarrow$ $\|y - \hat{y}\|$ is minimized

- What do $\hat{y}$ and $\boldsymbol{e}$ look like?



Orthogonal Projections

Orthogonal Complement

- Consider a vector space V and a subspace W.

- We say that $x \perp W$ if $\forall\, w \in W \;\; (x, w) = 0$.

Orthogonal Complement

- Consider a vector space V and a subspace W.
- We say that $x \perp W$ if $\forall\, w \in W \;\; (x, w) = 0$.
- <u>Example:</u>

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \perp span\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}$$

girafe ai

Orthogonal Complement

- Consider a vector space V and a subspace W.

- Orthogonal complement of W:

$$W_{\perp} = \{x \in V \mid (x, w) = 0 \ \forall w \in W\}$$

Orthogonal Complement

- Consider a vector space V and a subspace W.
- Orthogonal complement of W:

$$W_{\perp} = \{x \in V \mid (x, w) = 0 \ \forall w \in W\}$$

- <u>Example:</u>

$$\mathrm{W} = span\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}, \qquad \mathrm{W}_{\perp} = span\left\{\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

girafe ai

Orthogonal Decomposition

- Consider a vector space V and a subspace W.
- $x \in V$ and $x \notin W$.

Orthogonal Decomposition

- Consider a vector space V and a subspace W.
- $x \in V$ and $x \notin W$.

- x can be decomposed into a sum of two vectors:

$$x = x_W + x_{W^\perp}, \qquad x_W \in W, x_{W^\perp} \in W^\perp$$

girafe ai

Orthogonal Decomposition

- Consider a vector space V and a subspace W.
- $x \in V$ and $x \notin W$.

- x can be decomposed into a sum of two vectors:

$$x = x_W + x_{W^\perp}, \qquad x_W \in W, x_{W^\perp} \in W^\perp$$

- x_W – **orthogonal projection** of x onto W.

girafe ai

Orthogonal Decomposition

- Consider a vector space V and a subspace W.
- $x \in V$ and $x \notin W$.

- x can be decomposed into a sum of two vectors:

$$x = x_W + x_{W^\perp}, \qquad x_W \in W, x_{W^\perp} \in W^\perp$$

- x_W – **orthogonal projection** of x onto W.

- x_W is the closest vector to x in W.

girafe ai