Math Refresher for DS

Practical Session 3

girafe
ai

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.

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.

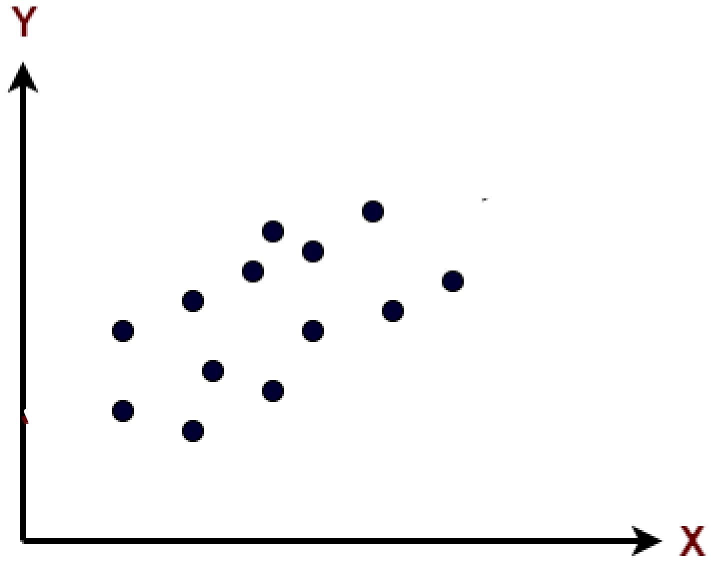
How to find a reasonable approximate solution?

Least Squares

Motivating Example

- Imagine that you have m observations:

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

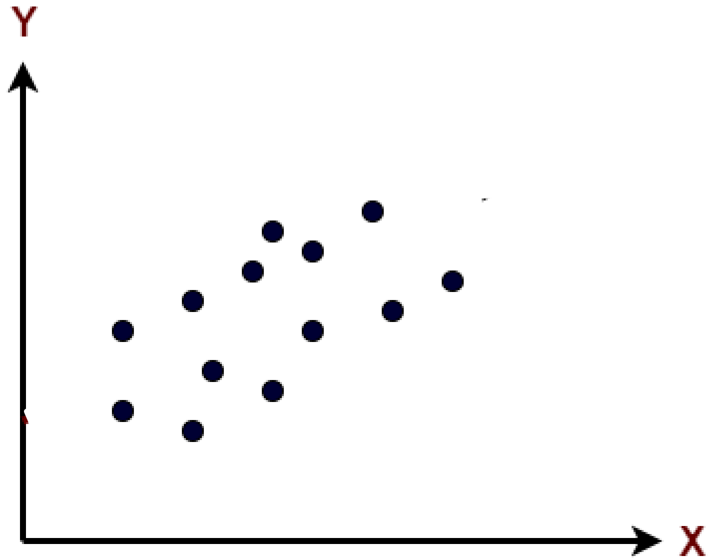


Motivating Example

- 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:

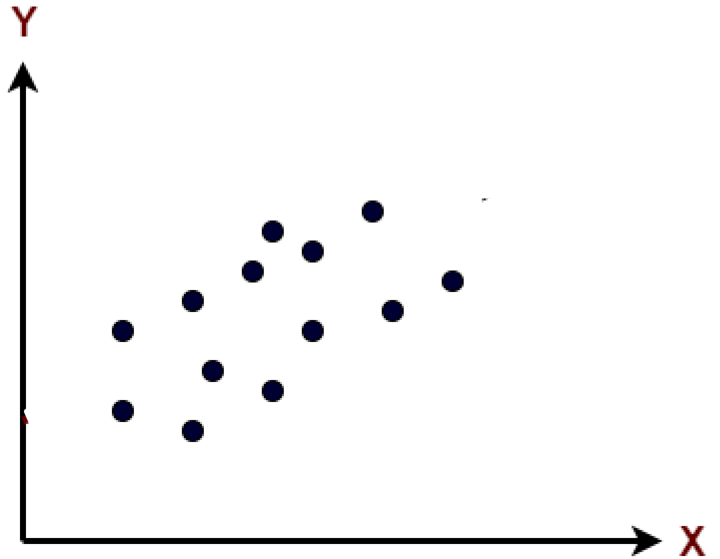


Motivating Example

- Imagine that you have m observations:

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

- You want to draw a line through your observations:



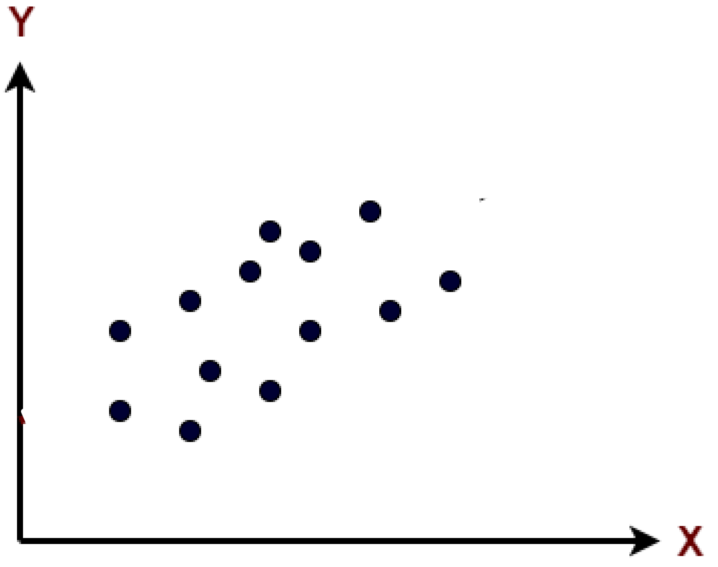
$$\begin{gathered} \boldsymbol{w_0} + \boldsymbol{w_1} x_1 = y_1 \\ \boldsymbol{w_0} + \boldsymbol{w_1} x_2 = y_n \\ \vdots \\ \boldsymbol{w_0} + \boldsymbol{w_1} x_m = y_m \end{gathered} \quad \Leftrightarrow$$

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:



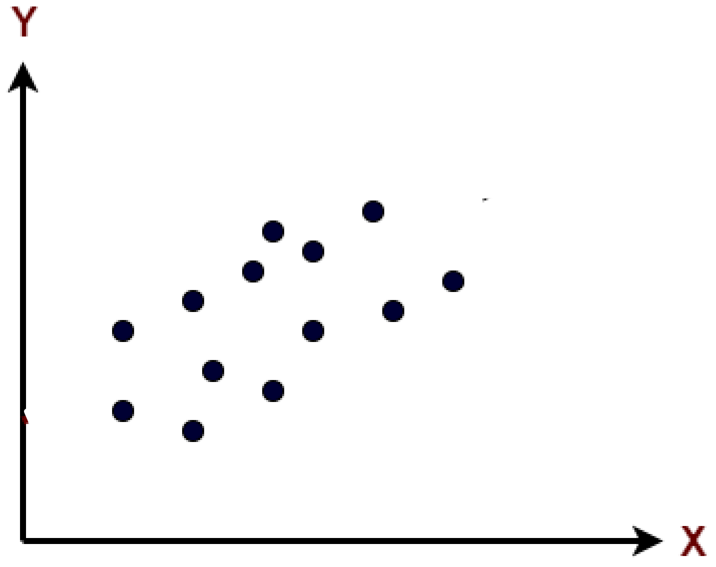
$$\begin{matrix} w_0 + w_1 x_1 = y_1 \\ w_0 + w_1 x_2 = y_n \\ \vdots \\ w_0 + w_1 x_m = y_m \end{matrix} \Leftrightarrow Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2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

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:



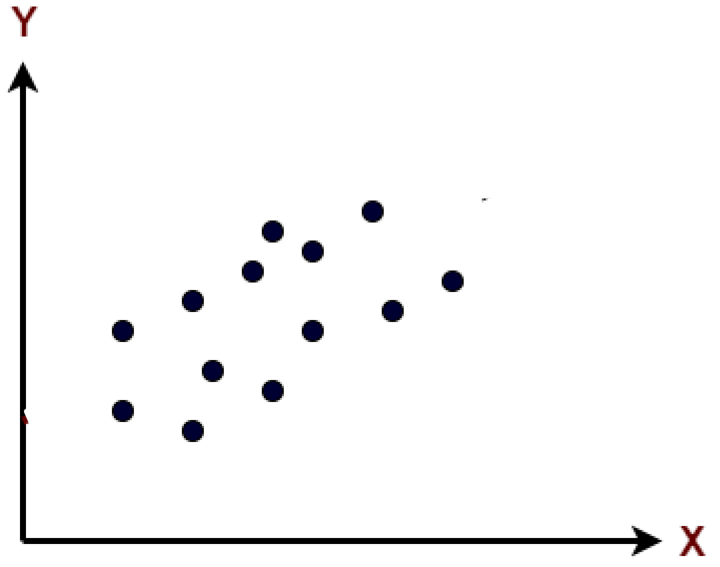
$$\begin{matrix} w_0 + w_1 x_1 = y_1 \\ w_0 + w_1 x_2 = y_n \\ \vdots \\ w_0 + w_1 x_m = y_m \end{matrix} \iff Xw = y, \text{ where } X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & m \end{pmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots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

- Overdetermined system, no solutions.

Motivating Example

- Imagine that you have m observations:

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



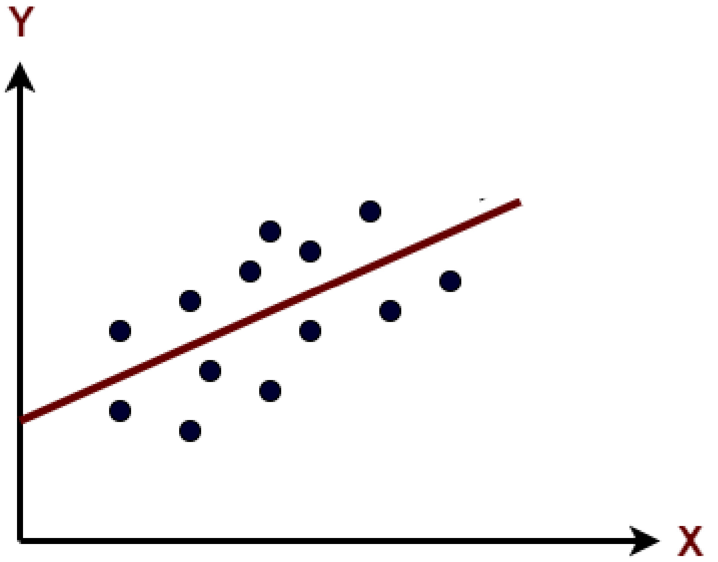
- You want to draw a line through your observations:

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

- Overdetermined system, no solutions.

- What shall we do?

Motivating Example



- Imagine that you have m observations:

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

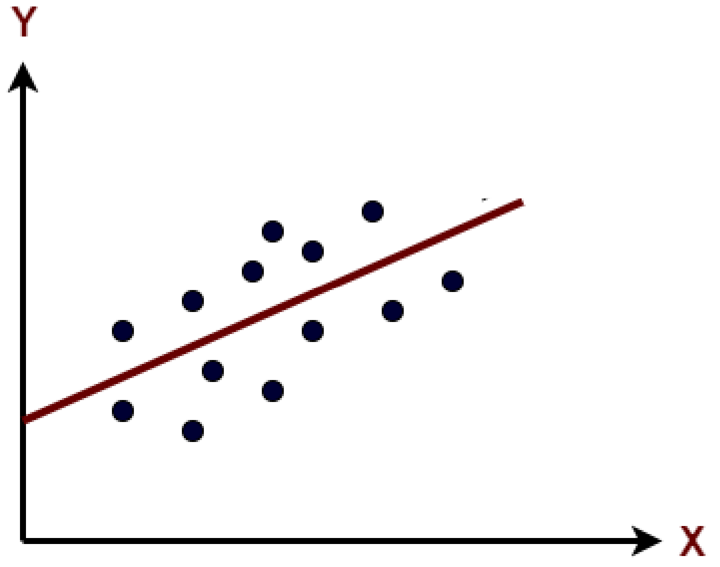
- You want to draw a line through your observations:

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

- Overdetermined system, no solutions.

- What shall we do?

Motivating Example



- Imagine that you have m observations:

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

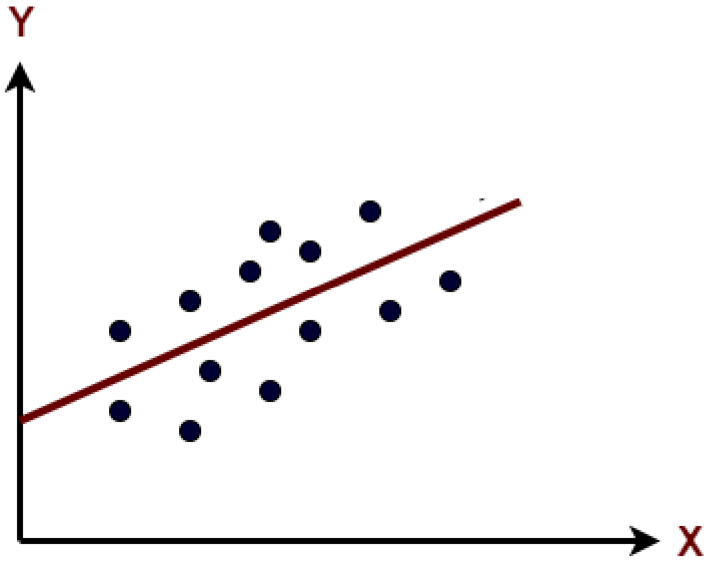
- You want to draw a line through your observations:

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

- 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:

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

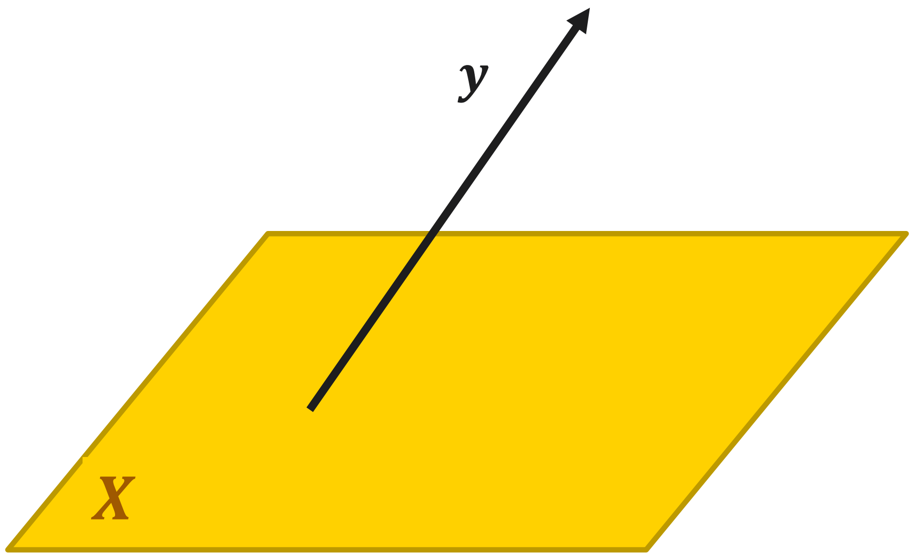
- You want to draw a line through your observations:

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

- 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.

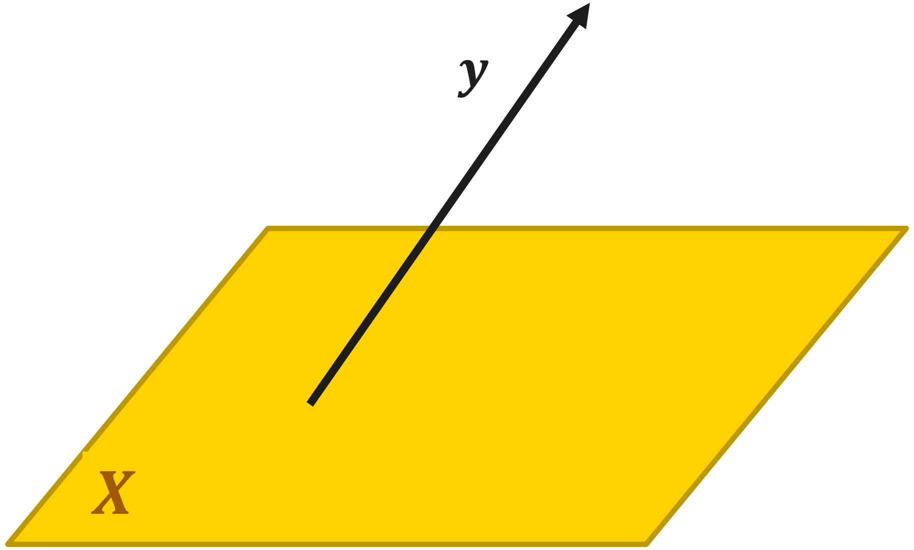
Method of Least Squares

- $Xw = y$ 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

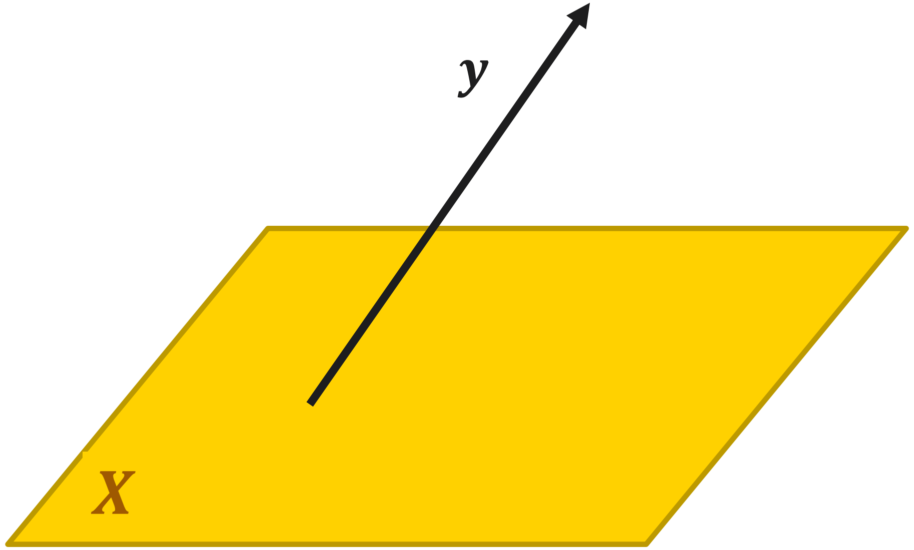


Method of Least Squares

- $Xw = y$ 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^*
  $\Leftrightarrow$



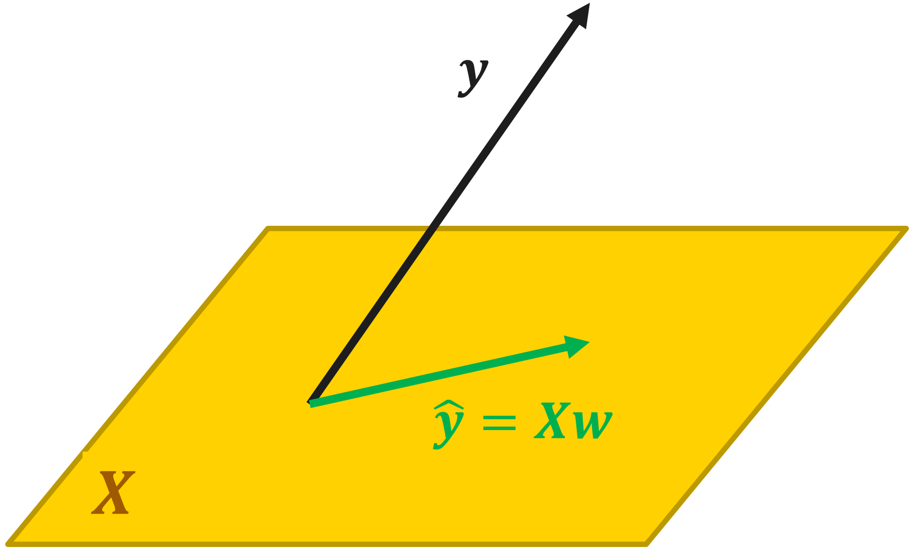
Method of Least Squares

- $Xw = y$ 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

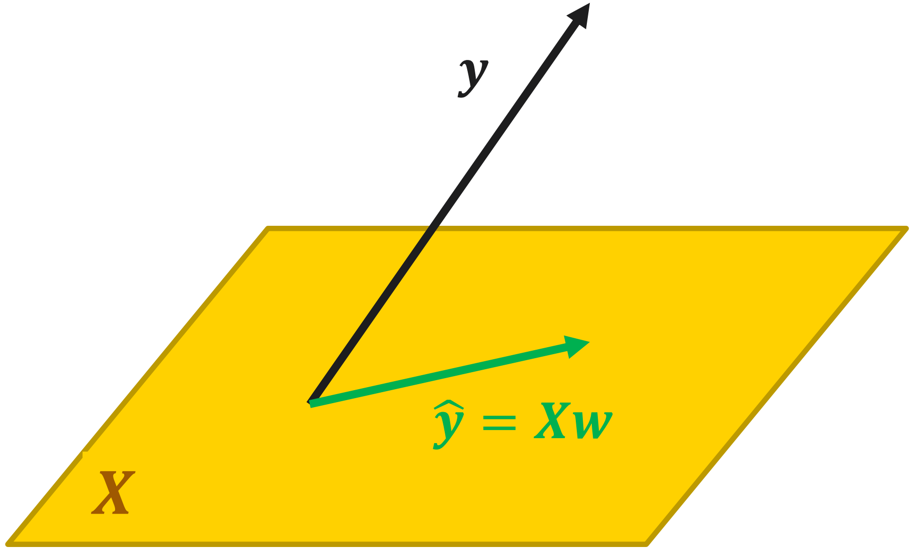
- $Xw = y$ has no solutions $\Leftrightarrow$ y is **not** in the column space of X.

- 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

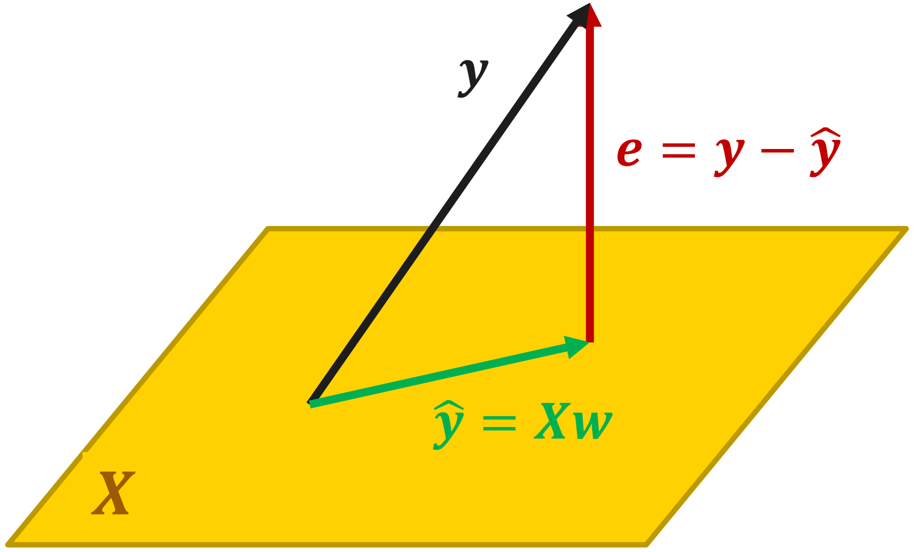
- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.

- 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.

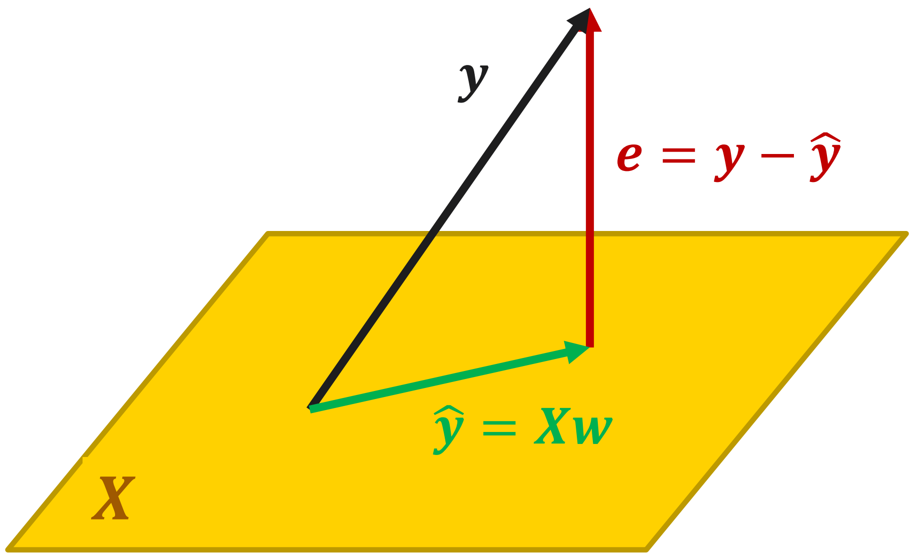
- Let's chose $\hat{y}$ such that

  $Xw = \hat{y}$ 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 $\forall\, w \in W \;\; (x, w) = 0$.
- Example:

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

Orthogonal Complement

- Consider a vector space V and a subspace W.

- Orthogonal complement of W:

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

girafe ai

Orthogonal Complement

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

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

- Example:

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

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:

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

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.

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

Method of Least Squares

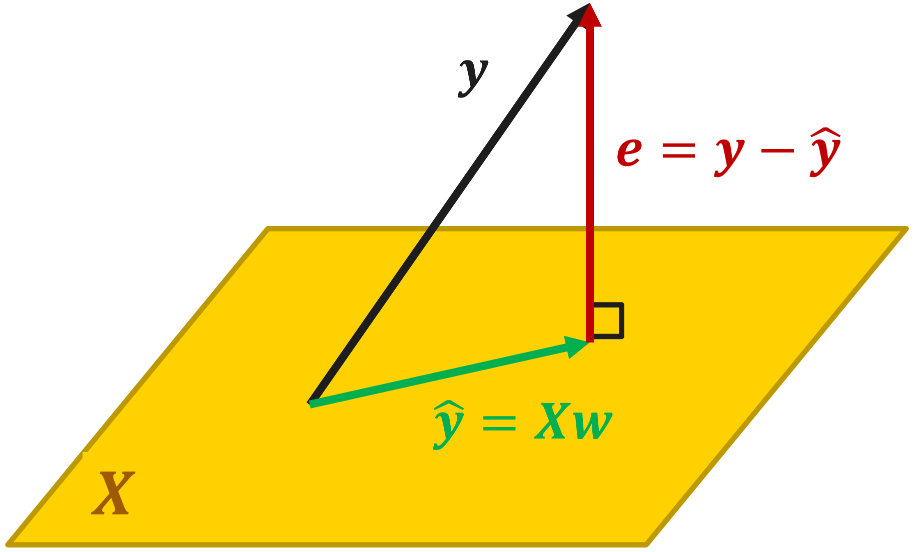
- Our goal: fit a hyperplane through the data $\left(\mathrm{x}^{i}, y\right)$.
- $Xw = y$ has no solutions $\Leftrightarrow$ y is **not** in the column space of X.

- 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?

Method of Least Squares

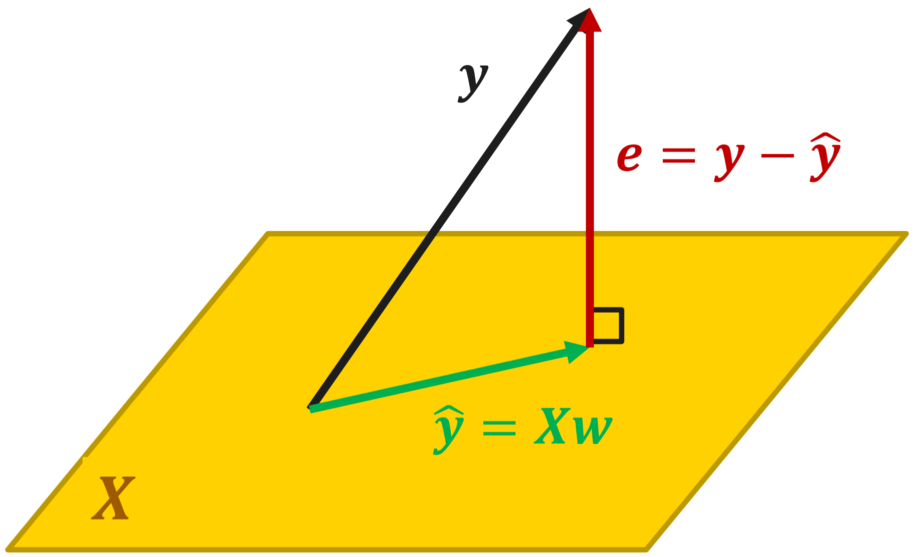
- Our goal: fit a hyperplane through the data (x^i, y).
- $Xw = y$ has no solutions $\Leftrightarrow$ y is **not** in the column space of X.

- 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?
  $\hat{y}$ is the orthogonal projection of $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$!

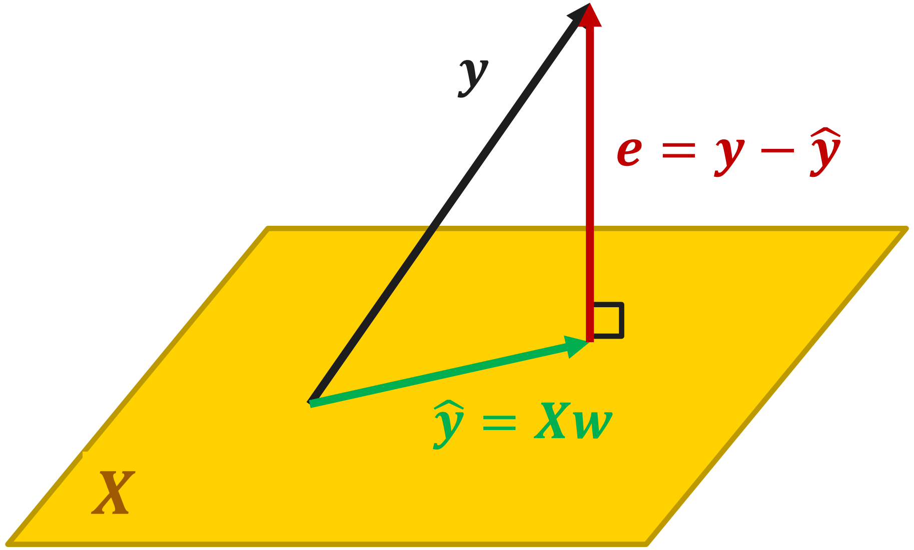
Method of Least Squares

- Our goal: fit a hyperplane through the data (x^i, y).
- $Xw = y$ 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

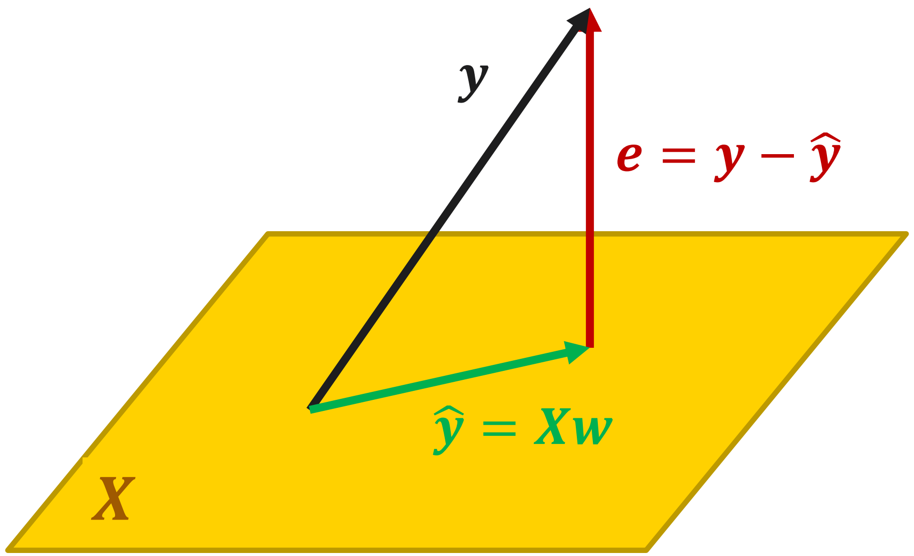


- What do $\hat{y}$ and $\boldsymbol{e}$ look like?
  $\hat{y}$ is the orthogonal projection of $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$!
  $\boldsymbol{e}$ is orthogonal onto the column space of $\boldsymbol{X}$.

Method of Least Squares

$$Xw^* = \hat{y} = y - e$$

$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

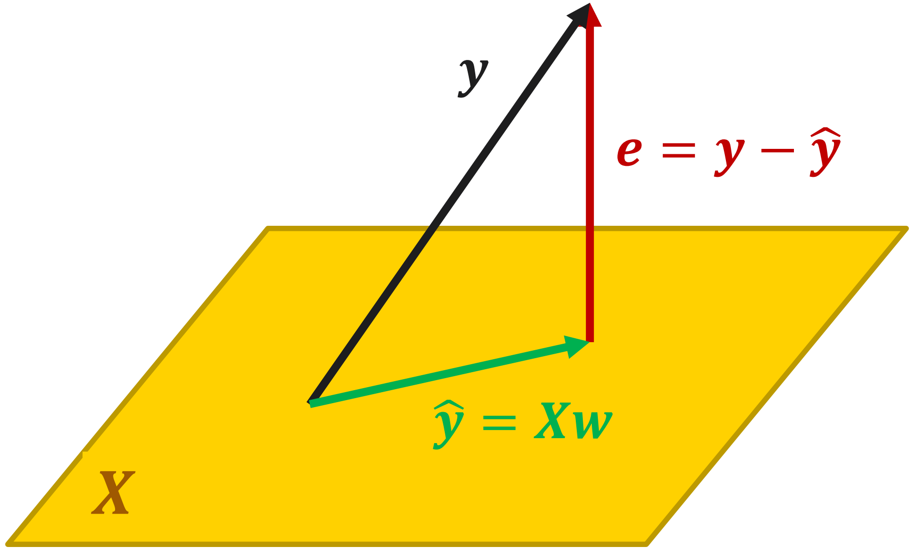


Method of Least Squares

$$\boldsymbol{X}\boldsymbol{w}^{*} = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

$$\boldsymbol{X}^{\boldsymbol{T}} X w^{*} = \boldsymbol{X}^{\boldsymbol{T}} y - \boldsymbol{X}^{\boldsymbol{T}} e$$

$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^{*} = ?$ – optimal weights

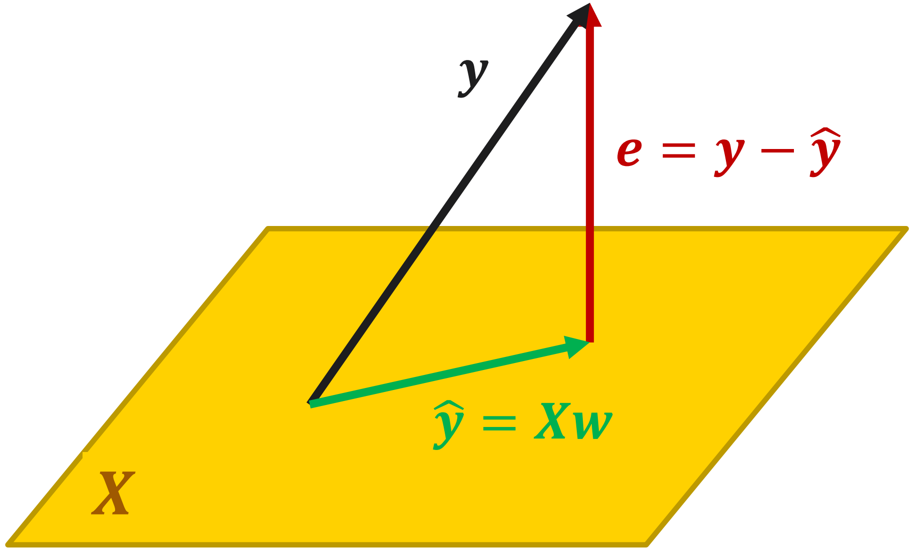


Method of Least Squares

$$\boldsymbol{Xw^*} = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

$$\boldsymbol{X^T} Xw^* = \boldsymbol{X^T} y - \boldsymbol{X^T} e$$

$\hat{y}$ - orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{col}(X)$
$w^* = ?$ – optimal weights



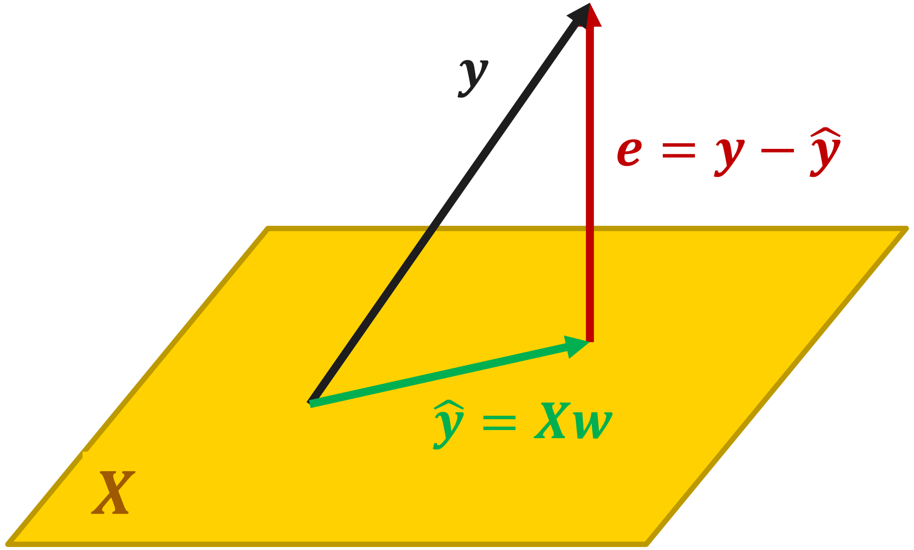
Method of Least Squares

$$\boldsymbol{Xw^*} = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

$$\boldsymbol{X^T} Xw^* = \boldsymbol{X^T} y - \boldsymbol{X^T} e$$

e is orthogonal to columns of $X \Rightarrow$



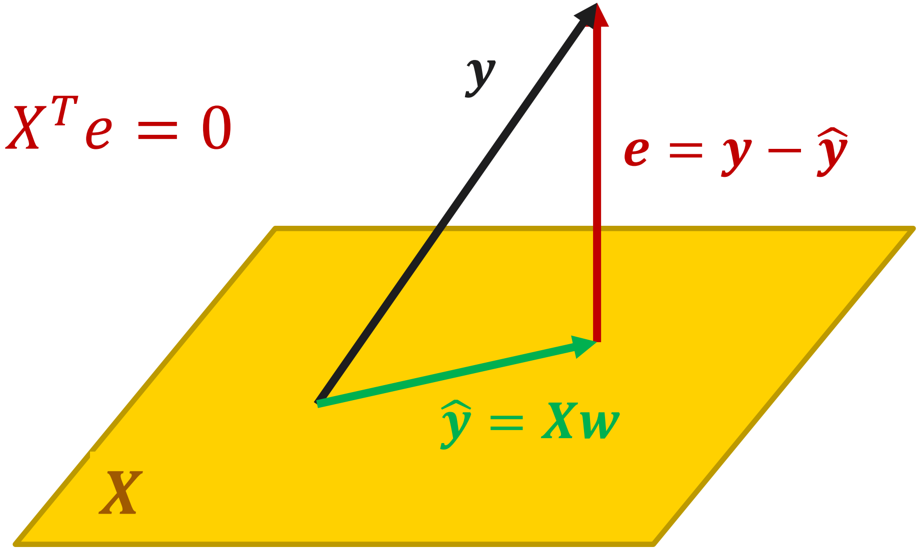
Method of Least Squares

$\boldsymbol{Xw}^* = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$

$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

$\boldsymbol{X^T} Xw^* = \boldsymbol{X^T} y - \boldsymbol{X^T} e$

e is orthogonal to columns of $X \Rightarrow X^T e = 0$



Method of Least Squares

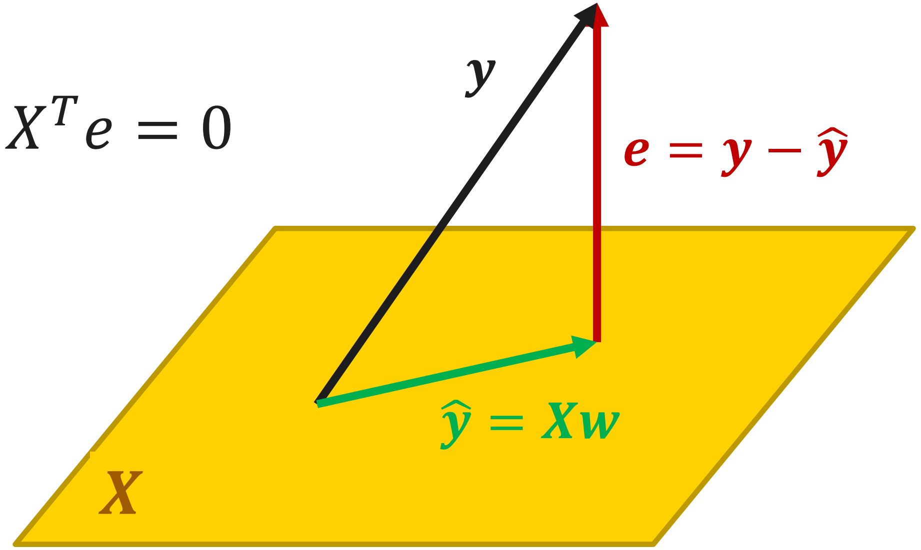
$$\boldsymbol{Xw}^* = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

$$\boldsymbol{X^T}Xw^* = \boldsymbol{X^T}y - \boldsymbol{X^T}e$$

e is orthogonal to columns of $X \Rightarrow X^T e = 0$

$$X^T X w^* = X^T y$$



Method of Least Squares

$$\boldsymbol{Xw^*} = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

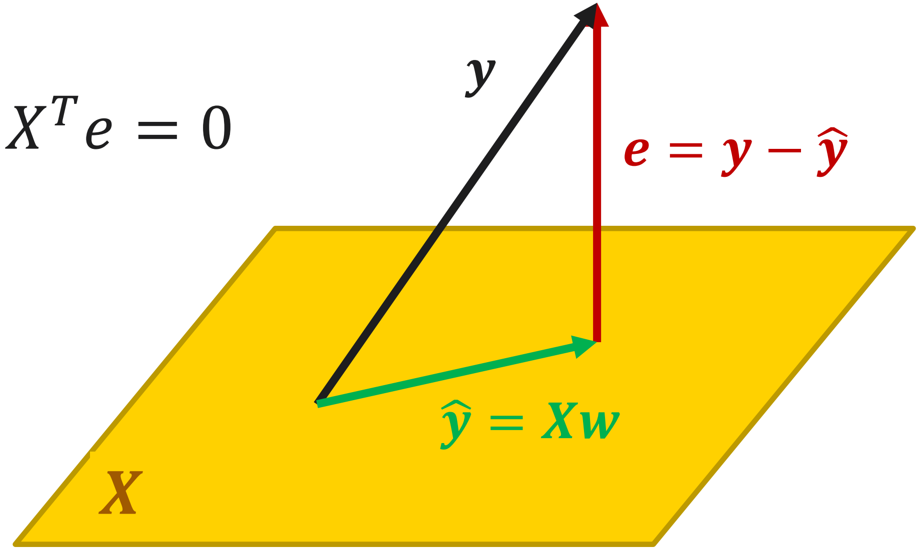
$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

$$\boldsymbol{X^T}Xw^* = \boldsymbol{X^T}y - \boldsymbol{X^T}e$$

e is orthogonal to columns of $X \Rightarrow X^T e = 0$

$$X^T X w^* = X^T y$$

$$\boldsymbol{w^*} =$$



Method of Least Squares

$$\boldsymbol{Xw^*} = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

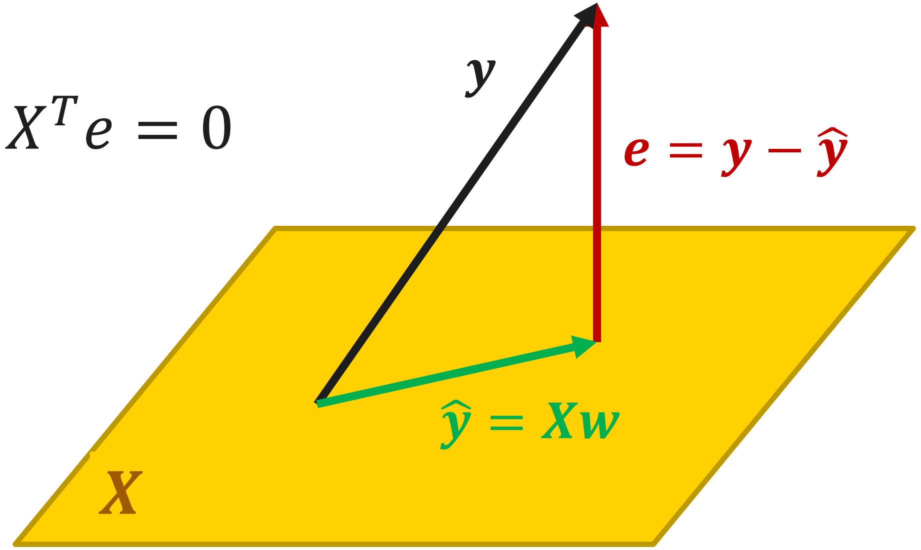
$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

$$\boldsymbol{X^T} X w^* = \boldsymbol{X^T} y - \boldsymbol{X^T} e$$

e is orthogonal to columns of $X \Rightarrow X^T e = 0$

$$X^T X w^* = X^T y$$

$\boldsymbol{w^* = (X^T X)^{-1} X^T y}$ – unknown coefficients.



Method of Least Squares

$$\boldsymbol{X}\boldsymbol{w}^* = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

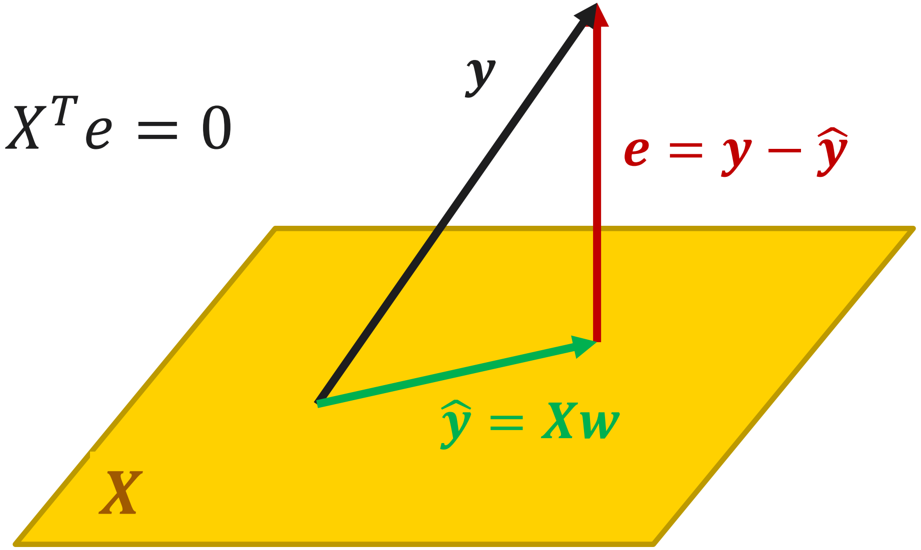
$\hat{y}$ - orthogonal projection of y onto $\boldsymbol{col}(X)$
$w^* = ?$ – optimal weights

$$\boldsymbol{X}^{\boldsymbol{T}} X w^* = \boldsymbol{X}^{\boldsymbol{T}} y - \boldsymbol{X}^{\boldsymbol{T}} e$$

e is orthogonal to columns of $X \Rightarrow X^T e = 0$

$$X^T X w^* = X^T y$$

$\boldsymbol{w}^* = \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$ – unknown coefficients.



$$\hat{\boldsymbol{y}} = X w^* =$$

Method of Least Squares

$$\boldsymbol{Xw}^* = \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{e}$$

$\hat{y}$ - orthogonal projection of y onto $col(X)$

$w^* = ?$ – optimal weights

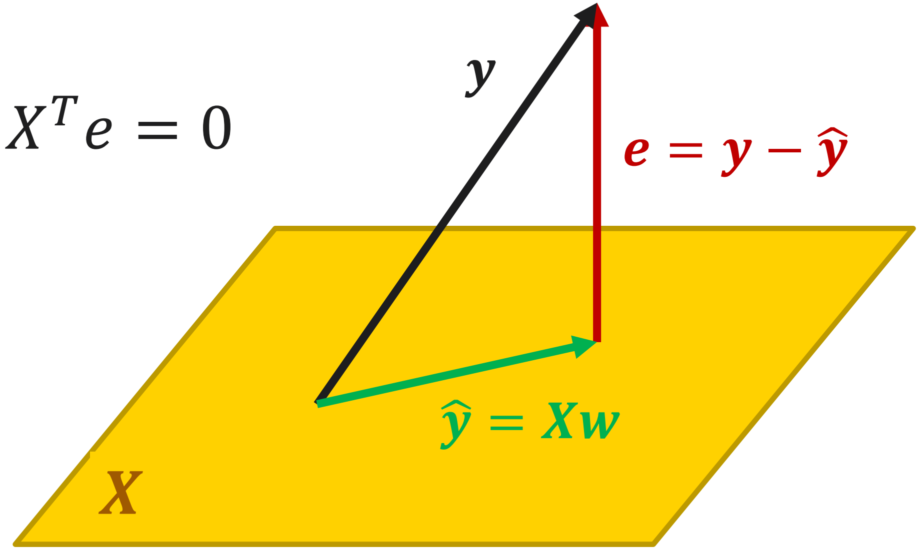
$$\boldsymbol{X}^{\boldsymbol{T}} Xw^* = \boldsymbol{X}^{\boldsymbol{T}} y - \boldsymbol{X}^{\boldsymbol{T}} e$$

e is orthogonal to columns of $X \Rightarrow X^T e = 0$

$$X^T X w^* = X^T y$$

$\boldsymbol{w}^* = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$ – unknown coefficients.

$$\hat{\boldsymbol{y}} = Xw^* = \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$$



Method of Least Squares

$$\boldsymbol{Xw}^* = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

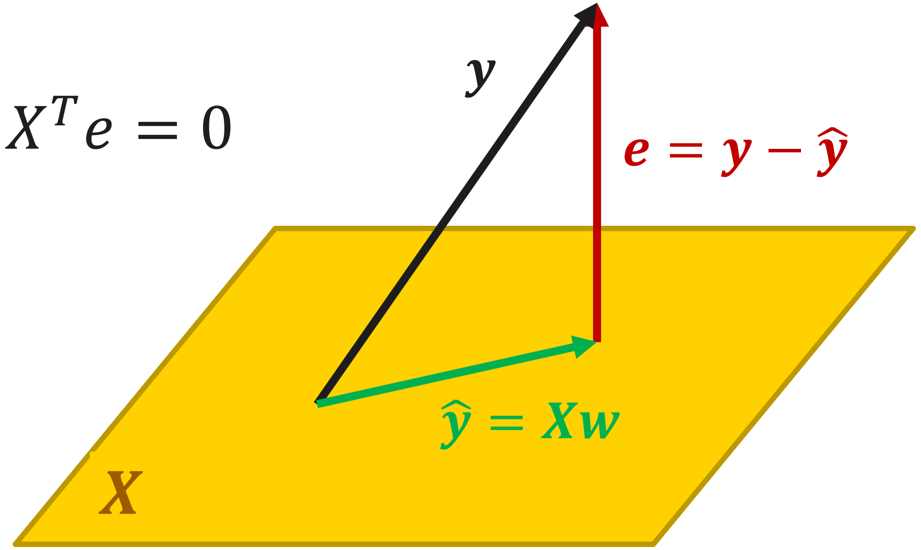
$$\boldsymbol{X^T}Xw^* = \boldsymbol{X^T}y - \boldsymbol{X^T}e$$

e is orthogonal to columns of $X \Rightarrow X^T e = 0$

$$X^T X w^* = X^T y$$

$\boldsymbol{w}^* = \left(\boldsymbol{X^T X}\right)^{-1}\boldsymbol{X^T y}$ – unknown coefficients.

$$\hat{\boldsymbol{y}} = Xw^* = \underbrace{\boldsymbol{X}\left(\boldsymbol{X^T X}\right)^{-1}\boldsymbol{X^T}}_{\text{projection matrix}}\boldsymbol{y}$$



Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\Longrightarrow \exists v \neq 0: \;\; X^T X v = 0:$$

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\implies \exists v \neq 0: \; X^T X v = 0:$$

$$X^T X v = 0$$

girafe ai

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\Longrightarrow \exists v \neq 0: \; X^T X v = 0:$$

$$X^T X v = 0$$

$$v^T X^T X v = 0$$

girafe ai

Why Does $(X^TX)^{-1}$ Exist?

- We derived that $w = (X^TX)^{-1}X^Ty$. But can we be sure $(X^TX)^{-1}$ exists?
- Suppose it doesn't.

$$\Longrightarrow \exists v \neq 0: \ X^TXv = 0:$$

$$X^TXv = 0$$

$$v^TX^TXv = 0$$

$$(Xv)^T(Xv) = 0$$

girafe ai

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\Longrightarrow \exists v \neq 0: \; X^T X v = 0:$$

$$X^T X v = 0$$

$$v^T X^T X v = 0$$

$$(Xv)^T (Xv) = 0$$

$$(Xv, Xv) = \|Xv\|^2 = 0 \;\Leftrightarrow$$

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\Longrightarrow \exists v \neq 0\text{:}\ \ X^T X v = 0\text{:}$$

$$X^T X v = 0$$

$$v^T X^T X v = 0$$

$$(Xv)^T (Xv) = 0$$

$$(Xv, Xv) = \|Xv\|^2 = 0 \ \Leftrightarrow Xv = 0$$

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\Longrightarrow \exists v \neq 0: \; X^T X v = 0:$$

$$X^T X v = 0$$

$$v^T X^T X v = 0$$

$$(Xv)^T (Xv) = 0$$

$$(Xv, Xv) = \|Xv\|^2 = 0 \; \Leftrightarrow Xv = 0$$

But $v \neq 0$ and $rank(X) = n$ (no linearly dependent columns)

girafe ai

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\Longrightarrow \exists v \neq 0: \; X^T X v = 0:$$

$$X^T X v = 0$$

$$v^T X^T X v = 0$$

$$(Xv)^T (Xv) = 0$$

$$(Xv, Xv) = \|Xv\|^2 = 0 \;\Leftrightarrow Xv = 0$$

But $v \neq 0$ and $rank(X) = n$ (no linearly dependent columns)

$\Longrightarrow$ contradiction.

girafe ai

Why Does $(X^T X)^{-1}$ Exist?

- We derived that $w = (X^T X)^{-1} X^T y$. But can we be sure $(X^T X)^{-1}$ exists?
- Suppose it doesn't.

$$\Rightarrow \exists v \neq 0:\ \ X^T X v = 0:$$

$$X^T X v = 0$$

$$v^T X^T X v = 0$$

$$(Xv)^T (Xv) = 0$$

$$(Xv, Xv) = \|Xv\|^2 = 0 \ \Leftrightarrow Xv = 0$$

But $v \neq 0$ and $rank(X) = n$ (no linearly dependent columns)

$\Rightarrow$ contradiction.

So, $(X^T X)^{-1}$ exists.

girafe ai

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

  Reminder: $w^* = (X^T X)^{-1} X^T y$

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

  Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

  Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \text{ – no exact solutions.}$$

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$ – no exact solutions.

$X^T X =$, $(X^T X)^{-1} =$, $(X^T X)^{-1} X^T =$

$w =$

girafe ai

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

  Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$ – no exact solutions.

$$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \qquad (X^T X)^{-1} = \qquad\qquad , \qquad (X^T X)^{-1} X^T =$$

$$w =$$

girafe ai

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$ – no exact solutions.

$$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \qquad (X^T X)^{-1} = \begin{bmatrix} 7/3 & -1 \\ -3/2 & 1 \end{bmatrix}, \qquad (X^T X)^{-1} X^T =$$

$$w =$$

girafe ai

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$ – no exact solutions.

$$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \qquad (X^T X)^{-1} = \begin{bmatrix} 7/3 & -1 \\ -3/2 & 1 \end{bmatrix}, \qquad (X^T X)^{-1} X^T = \begin{bmatrix} 4/3 & 1/3 & -2/3 \\ -1/2 & 0 & 1/2 \end{bmatrix}$$

$$w =$$

girafe ai

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

  Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \text{ – no exact solutions.}$$

$$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \qquad (X^T X)^{-1} = \begin{bmatrix} 7/3 & -1 \\ -3/2 & 1 \end{bmatrix}, \qquad (X^T X)^{-1} X^T = \begin{bmatrix} 4/3 & 1/3 & -2/3 \\ -1/2 & 0 & 1/2 \end{bmatrix}$$

$$w = (X^T X)^{-1} X^T y = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \rightarrow \text{regression line}$$

Toy Example

- Observations (x_i, y_i):

$$(1,1), \qquad (2,3), \qquad (3,2)$$

- With least squares, fit a line $y = w_0 + w_1 x$ through these points.

  Reminder: $w^* = (X^T X)^{-1} X^T y$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$ – no exact solutions.

$$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \qquad (X^T X)^{-1} = \begin{bmatrix} 7/3 & -1 \\ -3/2 & 1 \end{bmatrix}, \qquad (X^T X)^{-1} X^T = \begin{bmatrix} 4/3 & 1/3 & -2/3 \\ -1/2 & 0 & 1/2 \end{bmatrix}$$

$$w = (X^T X)^{-1} X^T y = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$ → regression line $y = 1 + 0.5x$.