girafe ai

Math Refresher for DS

Practical Session 4

Today

- Least Squares (continued)
- More on coordinates change

Graded Assignment 1 is OUT

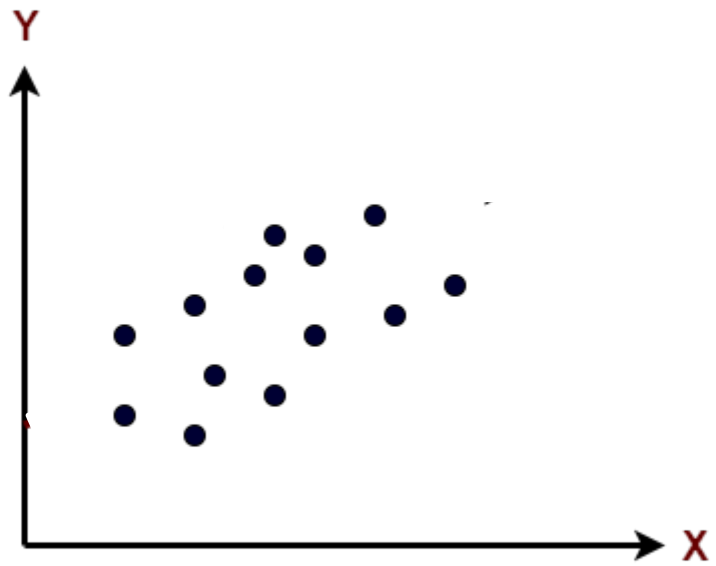
- Google-form, link in repo and on Telegram.
- Submit answers <u>and</u> detailed solutions.

- Submission deadline: Monday, October 17, 23:59 AoE
- Late submissions won’t be accepted.

Where we stopped last time…

Method of Least Squares

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



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.

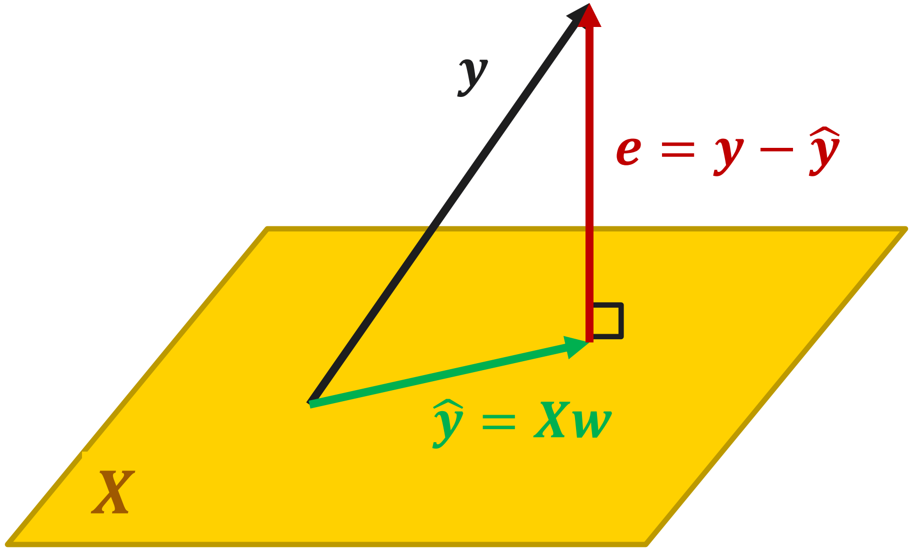
- Let's chose $\hat{y}$ such that

  $Xw = \hat{y}$ 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 e look like?



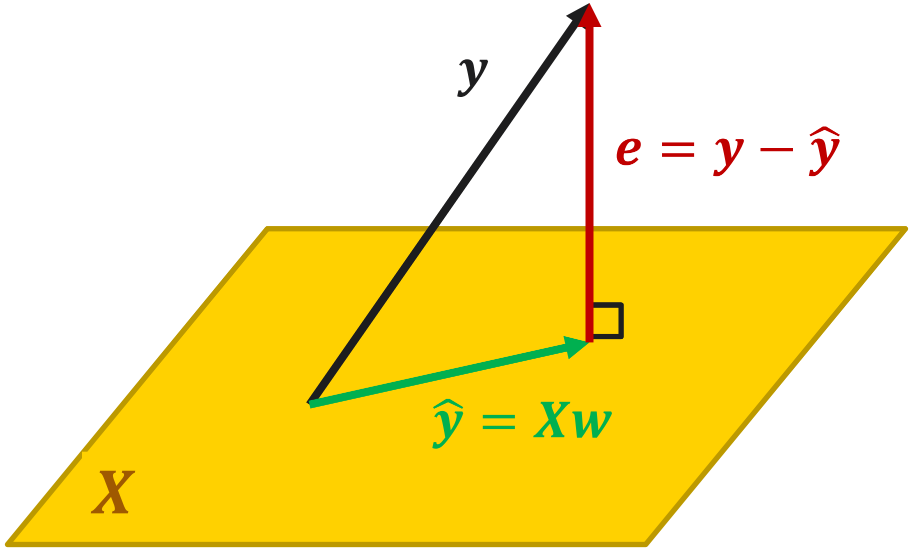
Method of Least Squares

- 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 $\boldsymbol{X}$!
  $\boldsymbol{e}$ is orthogonal onto the column space of $\boldsymbol{X}$.

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:

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

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:

$$\mathrm{V} = \mathbb{R}^3, \qquad \mathrm{W} = span\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}, \qquad \mathrm{W}_{\perp} = span\left\{\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

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

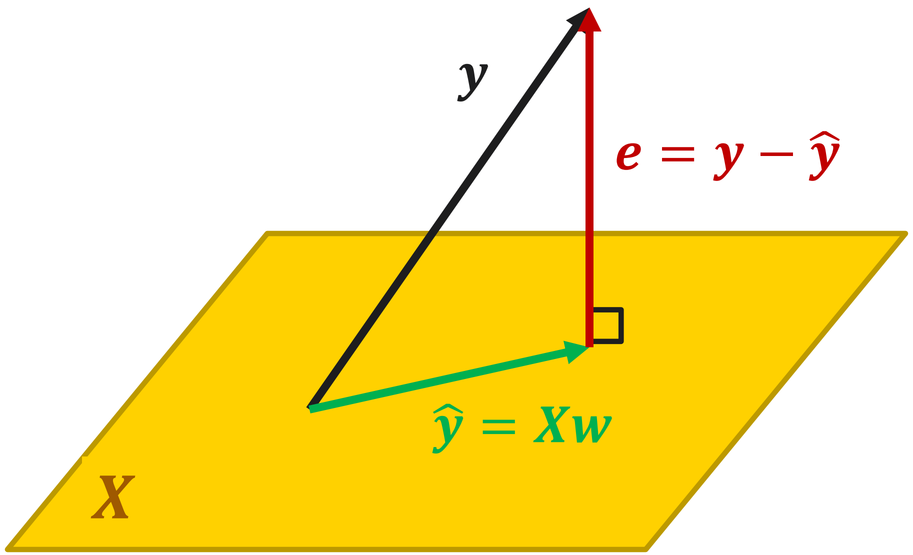
- Our goal: fit a hyperplane through the data (x^i, y).
- $Xw = y$ has no solutions $\Leftrightarrow$ y is **not** in the column space of X.

- 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?

Method of Least Squares

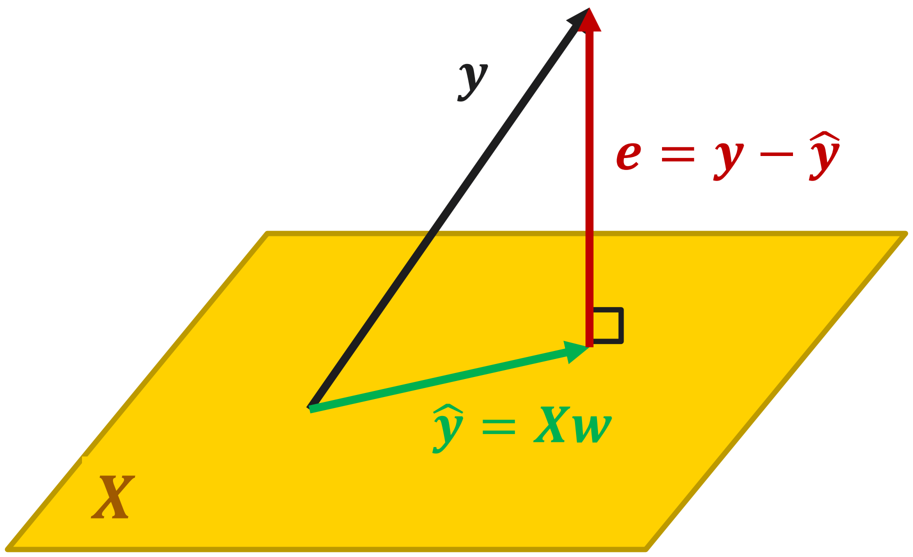
- Our goal: fit a hyperplane through the data (x^i, y).
- $Xw = y$ has no solutions $\Leftrightarrow$ y is not in the column space of X.

- 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 $\boldsymbol{X}$!

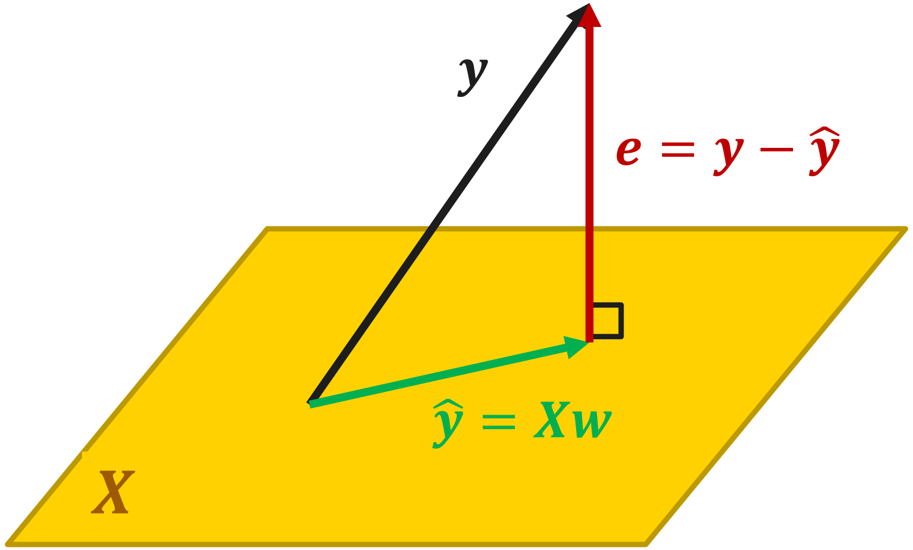
Method of Least Squares

- 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

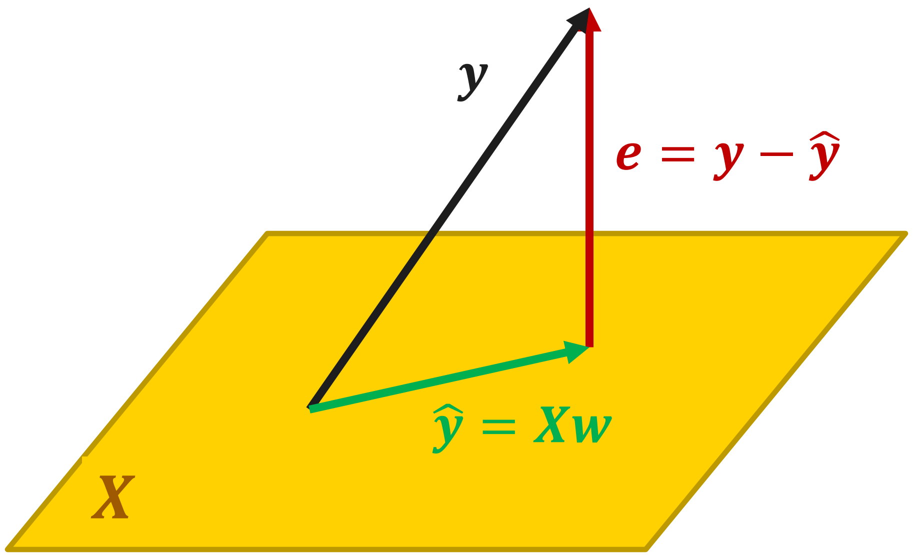


- What do $\hat{y}$ and $\boldsymbol{e}$ look like?
  $\hat{y}$ is the orthogonal projection of $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$!
  $\boldsymbol{e}$ is orthogonal onto the column space of $\boldsymbol{X}$.

Method of Least Squares

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

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

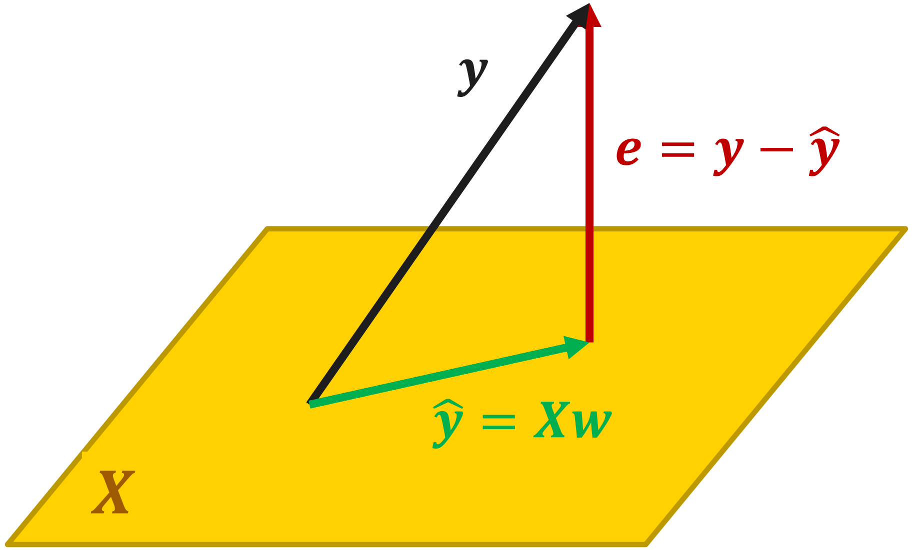


Method of Least Squares

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

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

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

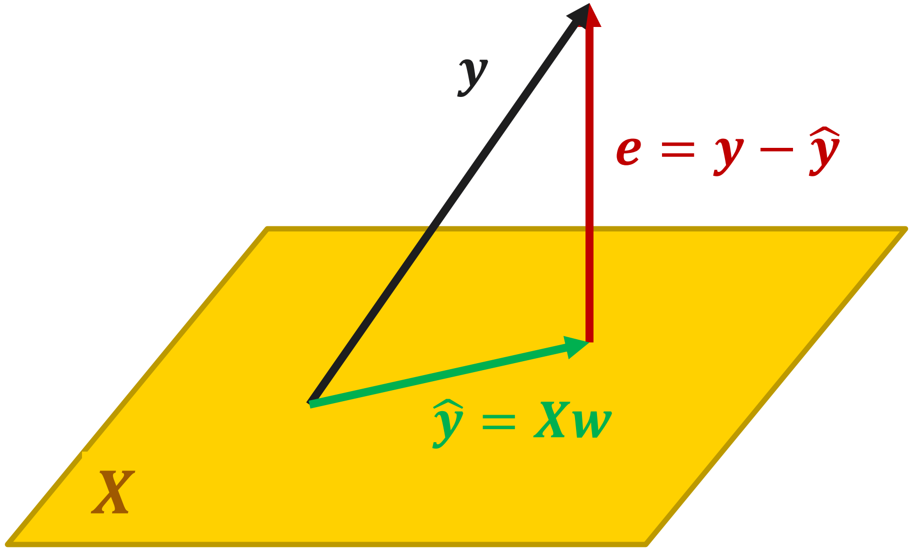


Method of Least Squares

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

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

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



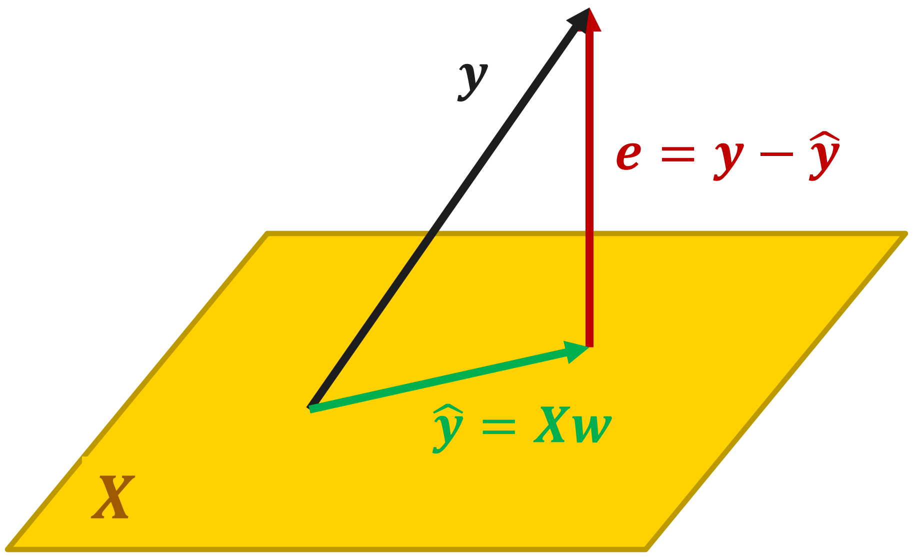
Method of Least Squares

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

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

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

e is orthogonal to columns of $X \Rightarrow$



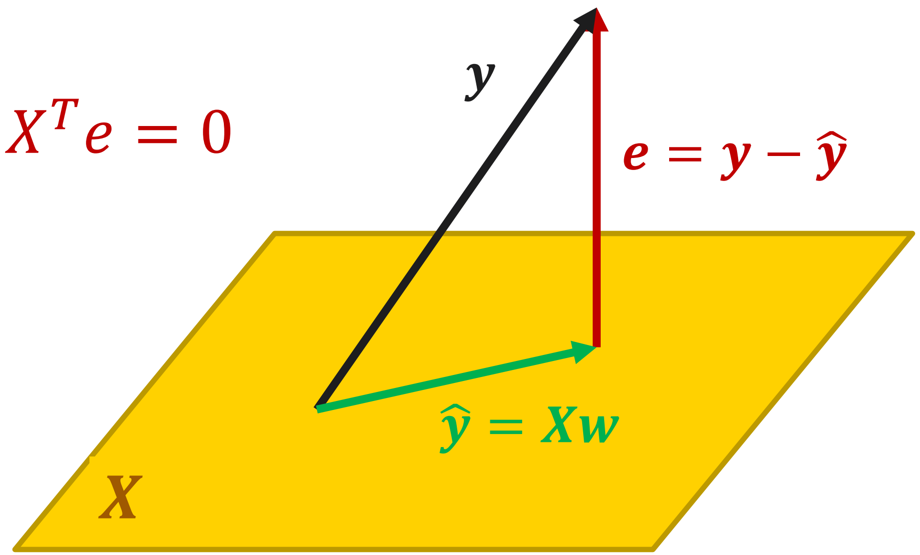
Method of Least Squares

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

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

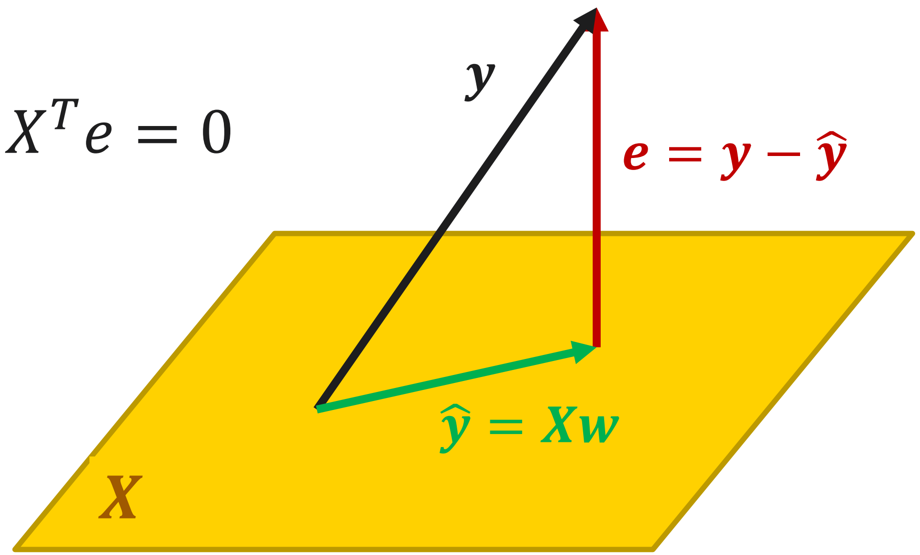
$$\boldsymbol{Xw}^* = \hat{y} = \boldsymbol{y} - \boldsymbol{e}$$

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

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

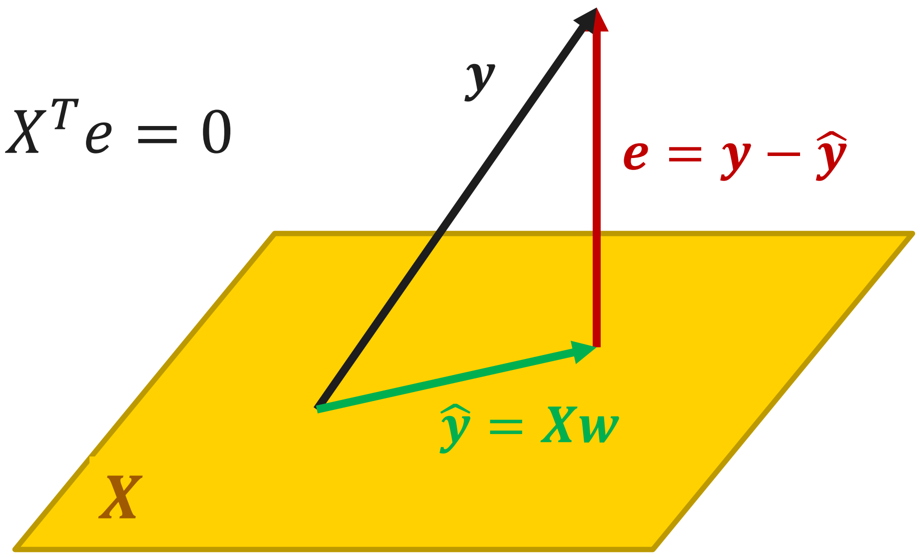
$\hat{y}$ - orthogonal projection of y onto $col(X)$
$w^* = ?$ – optimal weights

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

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

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

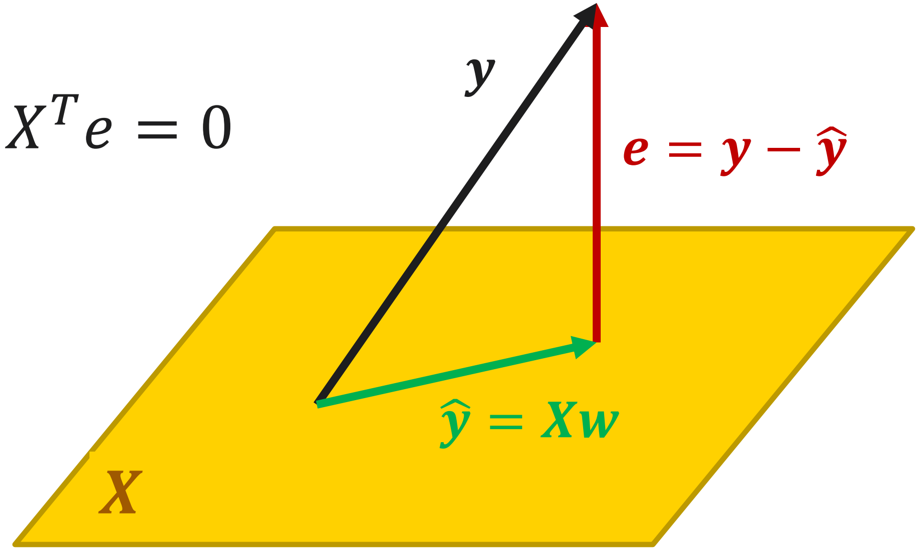
$\hat{y}$ - orthogonal projection of y onto $\boldsymbol{col}(X)$
$w^* = ?$ – optimal weights

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

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.



Method of Least Squares

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

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

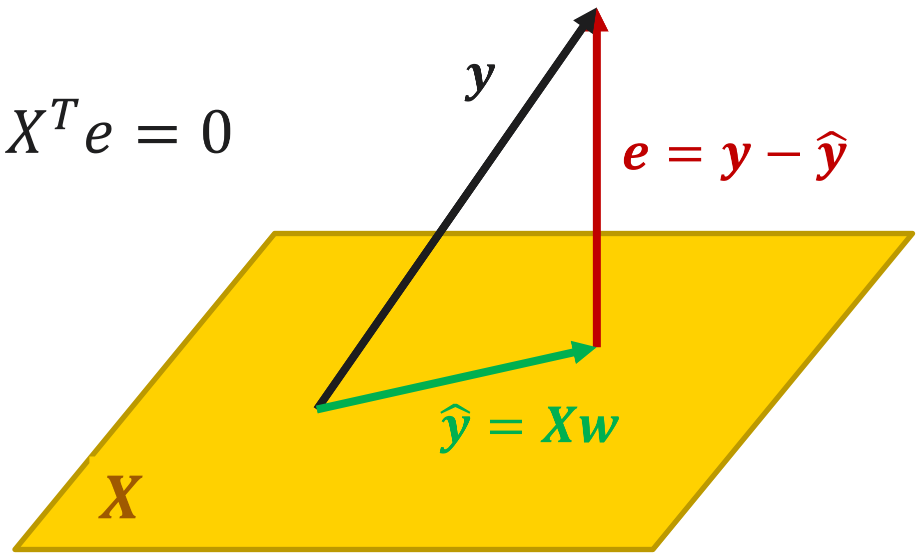
$$\boldsymbol{X}^{\boldsymbol{T}} Xw^* = \boldsymbol{X}^{\boldsymbol{T}} y - \boldsymbol{X}^{\boldsymbol{T}} e$$

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

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

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

$$\widehat{\boldsymbol{y}} = Xw^* =$$



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

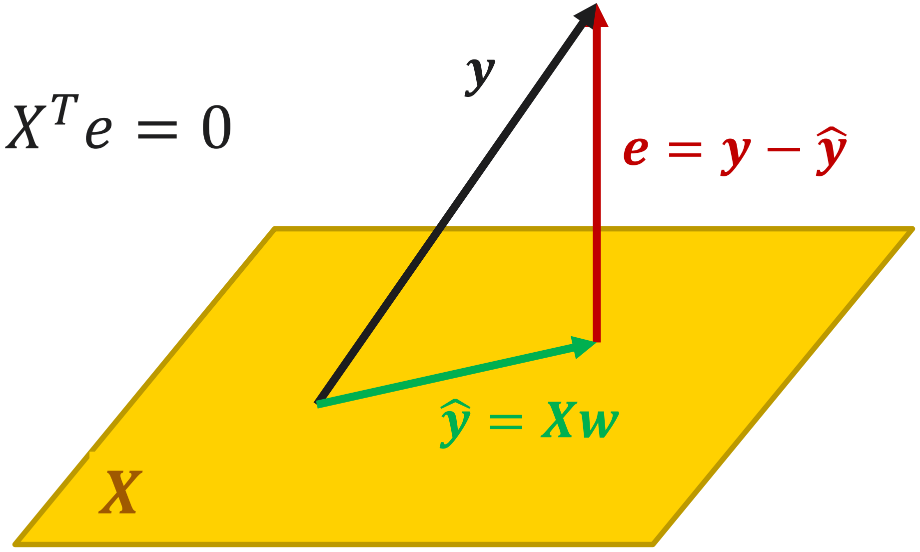
$$\boldsymbol{X^T} Xw^* = \boldsymbol{X^T} y - \boldsymbol{X^T} e$$

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

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

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

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



Method of Least Squares

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

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

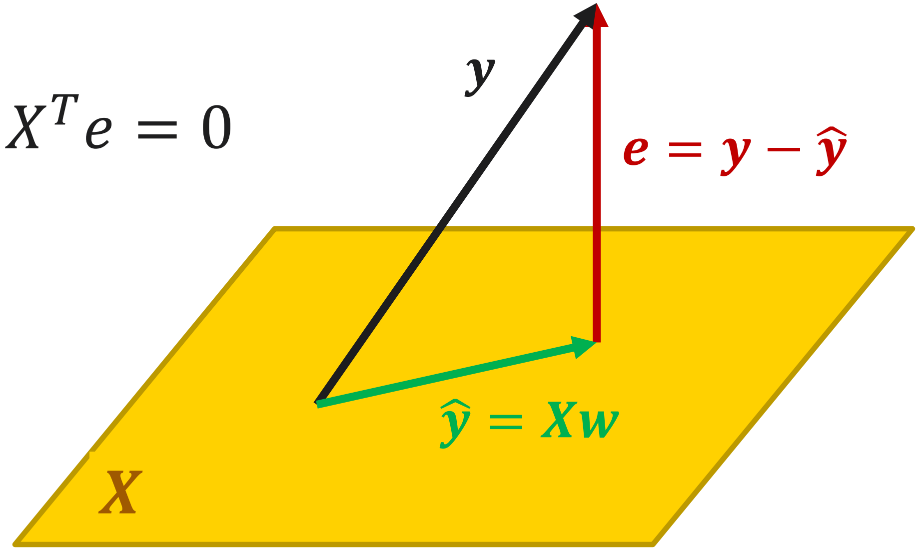
$$\boldsymbol{X}^{\boldsymbol{T}} X w^* = \boldsymbol{X}^{\boldsymbol{T}} y - \boldsymbol{X}^{\boldsymbol{T}} e$$

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

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

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

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



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.

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

$$X^T X v = 0$$

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.

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

$$X^T X v = 0$$

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

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.

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

$$X^T X v = 0$$

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

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

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.

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

$$X^T X v = 0$$

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

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

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

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.

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

$$X^T X v = 0$$

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

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

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

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

$\Longrightarrow$ contradiction.

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.

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

$$X^TXv = 0$$

$$v^TX^TXv = 0$$

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

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

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

$\Longrightarrow$ contradiction.

So, $(X^TX)^{-1}$ 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: $w^* = (X^T X)^{-1} X^T y$

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

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

$w =$

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

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

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

$$w =$$

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

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

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

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

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.

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

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

Let’s practice more!

https://colab.research.google.com/drive/1v_dDH5aSx9pQG4SSCLzNjKuCk6rwNWzf?usp=sharing

girafe ai

Coordinates Change

Change of Basis for Vectors

- V – a vector space.
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis.

- $x \in V$ – some vector.

- We know already that

Change of Basis for Vectors

- V – a vector space.
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis.

- $x \in V$ – some vector.

- We know already that

$$x_B = Mx_S, \ M = M_{B \to S} = \big[[s_1]_B \mid \dots \mid [s_n]_B\big] \text{ – transition matrix.}$$

Change of Basis for Vectors

- V – a vector space.
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis.

- $x \in V$ – some vector.

- We know already that

$$x_B = Mx_S,\ \ M = M_{B\to S} = \left[[s_1]_B \mid \dots \mid [s_n]_B\right] \text{ – transition matrix.}$$

- We also established that

$$\exists M^{-1} = M_{S\to B}, \qquad x_S = M^{-1}x_B$$

Change of Basis for Vectors

- V – a vector space.
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis.

- $x \in V$ – some vector.

- We know already that

$$x_B = Mx_S, \; M = M_{B \to S} = \big[[s_1]_B \mid \dots \mid [s_n]_B\big] \text{ – transition matrix.}$$

- We also established that

$$\exists M^{-1} = M_{S \to B}, \qquad x_S = M^{-1} x_B$$

- But vectors aren’t the only things with coordinates…

girafe ai

Change of Basis for Linear Transforms

- Consider a linear transform A.
- It's defined by its matrix: columns = what happens to basis vectors.

- Example: rotation

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Change of Basis for Linear Transforms

- Consider a linear transform A.
- It's defined by its matrix: columns = what happens to basis vectors.

- Example: rotation

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

- $S = \left\{ s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ – another basis.

How would A look like in this basis?

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

girafe ai

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = \quad [\boldsymbol{A}]_{\boldsymbol{S}} \quad \cdot [x]_S$$

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = \quad \boldsymbol{[A]_S} \quad \cdot [x]_S$$

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = \qquad\qquad\qquad \cdot [x]_S$$

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = \qquad\qquad M \cdot [x]_S$$

girafe ai

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = \qquad [A]_B M \cdot [x]_S$$

girafe ai

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = M^{-1}[A]_B M \cdot [x]_S$$

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = M^{-1}[A]_B M \cdot [x]_S$$

girafe ai

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_s = M^{-1}[A]_B M \cdot [x]_S$$

$$[\boldsymbol{A}]_{\boldsymbol{S}} =$$

Change of Basis for Linear Transforms

- A – linear transform;
- $B = \{b_1, \dots, b_n\}$ – current basis, $S = \{s_1, \dots, s_n\}$ – new basis;
- $M = M_{B \to S}$ – transition matrix.

- x – some vector.

$$[x']_B = [A]_B \cdot [x]_B$$

$$[x']_S = M^{-1}[A]_B M \cdot [x]_S$$

$$\mathbf{[A]_S = M^{-1}[A]_B M}$$

Change of Basis for Linear Transforms

- Back to our example:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad S\left\{s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\} -$$ new basis.

$$[A]_S = M^{-1}AM =$$

Change of Basis for Linear Transforms

- Back to our example:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad S\left\{s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\} - \text{ new basis.}$$

$$[A]_S = M^{-1}AM = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} =$$

Change of Basis for Linear Transforms

- Back to our example:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad S\left\{s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\} -$$ new basis.

$$[A]_S = M^{-1}AM = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} =$$

Change of Basis for Linear Transforms

- Back to our example:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad S\left\{s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\} - \text{ new basis.}$$

$$[A]_S = M^{-1}AM = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1/3 & -1/3 \\ 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} =$$

Change of Basis for Linear Transforms

- Back to our example:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad S\left\{s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\} - \text{ new basis.}$$

$$[A]_S = M^{-1}AM = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1/3 & -1/3 \\ 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix}$$

Change of Basis for Linear Transforms

$[A]_E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $[A]_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix}$, $S = \left\{ s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ – new basis;

$$x_S = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad x'_S = [A]_S \cdot x_S =$$

Change of Basis for Linear Transforms

$$[A]_E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},\ [A]_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix},\ S = \left\{ s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$ – new basis;

$$x_S = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad x'_S = [A]_S \cdot x_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} =$$

Change of Basis for Linear Transforms

$$[A]_E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},\ [A]_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix},\ S = \left\{ s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$ – new basis;

$$x_S = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad x'_S = [A]_S \cdot x_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Change of Basis for Linear Transforms

$[A]_E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $[A]_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix}$, $S = \left\{ s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ – new basis;

$$x_S = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad x'_S = [A]_S \cdot x_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$x_E =$$

girafe ai

Change of Basis for Linear Transforms

$[A]_E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $[A]_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix}$, $S = \left\{ s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ – new basis;

$$x_S = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad x'_S = [A]_S \cdot x_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$x_E = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \qquad x'_E =$$

girafe ai

Change of Basis for Linear Transforms

$[A]_E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $[A]_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix}$, $S = \left\{ s_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ – new basis;

$$x_S = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad x'_S = [A]_S \cdot x_S = \begin{bmatrix} 1/3 & -2/3 \\ 5/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$x_E = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \qquad x'_E = [A]_E \cdot x_E = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$$

Change of Basis for Linear Transforms

- Another example:

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}, \qquad S = \left\{ s_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} - \text{ new basis}$$

$$[A]_S = M^{-1}AM =$$

Change of Basis for Linear Transforms

- Another example:

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}, \qquad S = \left\{ s_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} - \text{ new basis}$$

$$[A]_S = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} =$$

Change of Basis for Linear Transforms

- Another example:

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}, \qquad S = \left\{ s_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} - \text{ new basis}$$

$$[A]_S = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} =$$

Change of Basis for Linear Transforms

- Another example:

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}, \qquad S = \left\{ s_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} -$$ new basis

$$[A]_S = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

Change of Basis for Linear Transforms

- Another example:

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}, \qquad S = \left\{ s_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, s_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} -$$ new basis

$$[A]_S = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

We get a diagonal matrix, it's easier to work with it!