EECS 245 Fall 2025
Math for ML

Lecture 16: Multiple Linear Regression

→ Read: Ch. 2.10 (new examples),
Ch. 3.1 (new), Ch. 3.2 (new)

Agenda

→ Recap: normal equations, design matrix, etc.

→ Multiple linear regression

→ Feature engineering

Chapter 3 and Chapter 2.10 → essential! actually read! it'll be useful for Midterm 2

Announcements

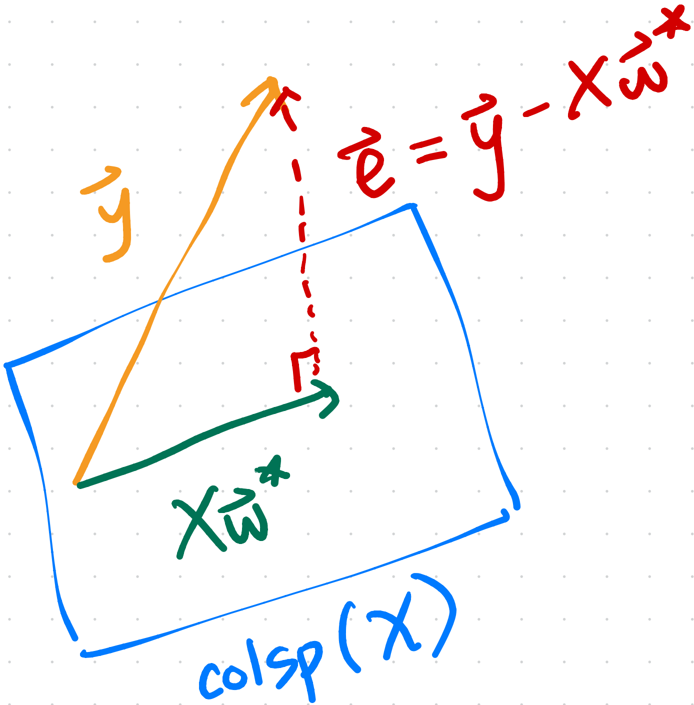
→ IA official app due today,

video + form due Saturday

→ HW 7 due tomorrow

→ HW 6 solutions out
→ read!

→ check Ed for research opportunities



$X : n \times d$

$\vec{y} : \mathbb{R}^n$

Goal: Find $\vec{w}^*$ that minimizes

$$\|\vec{y} - X\vec{w}\|^2$$

$\Rightarrow$ solution: pick $\vec{w}^*$ such that

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

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

"the normal equation"

2.10

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

① If X's columns are LI → X^TX invertible

review!

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

unique solution

$(AB)^{-1}$

$= B^{-1}A^{-1}$

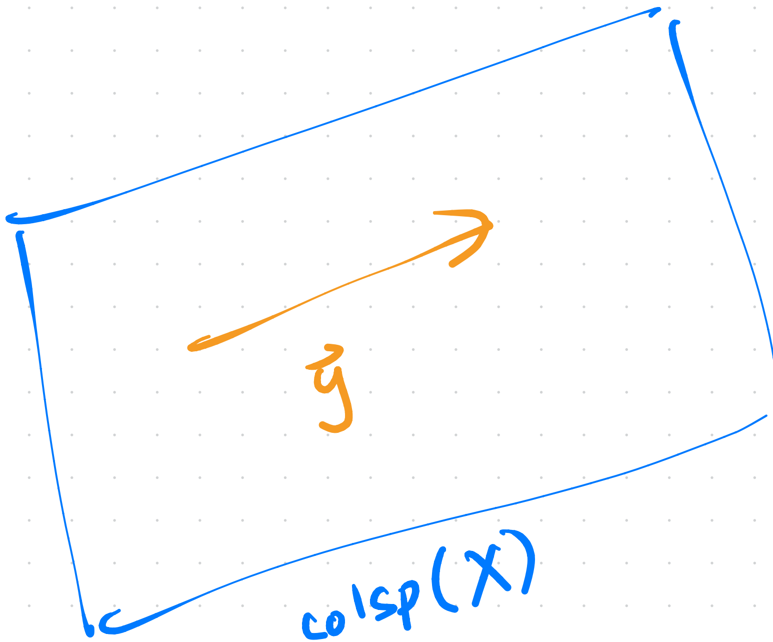
why can't I use this?

X isn't square!

② If X's columns aren't LI,

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

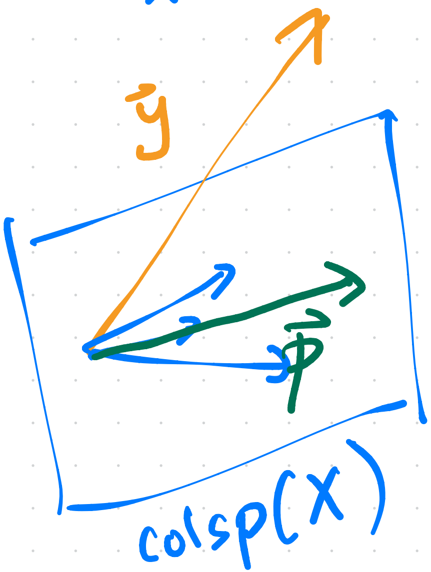
has infinitely many sol's $\vec{w}^*$



<u>if</u> X is square and has LI columns, this is the picture

(but usually, X <u>isn't</u> square)

$$X = \begin{array}{ccc} \begin{bmatrix} 3 & 1 & 0 \\ 6 & 2 & 1 \\ 3 & 1 & 0 \end{bmatrix} \\ \vec{x}^{(1)} \quad \vec{x}^{(2)} \quad \vec{x}^{(3)} \end{array} \qquad \vec{y} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \qquad X^T X \vec{w}^* = X^T \vec{y}$$



$\vec{p} = X\vec{w}^*$ is unique, but w^* isn't

$$2\vec{x}^{(1)} + \vec{x}^{(3)}$$
$$= X \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$
$$= \vec{x}^{(1)} + 3\vec{x}^{(2)} + \vec{x}^{(3)}$$
$$= X \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$

In Ch 2.8, we proved that

$$\text{nullsp}(X) = \text{nullsp}(X^T X)$$

... huh?

if $X\vec{v} = \vec{0}$, then $X^T X \vec{v} = \vec{0}$ and vice versa

suppose $\vec{w}'$ is a solution to

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

suppose $\vec{n} \in \text{nullsp}(X^T X)$ i.e. $X^T X \vec{n} = \vec{0}$

$$X^T X(\vec{w}' + \vec{n}) = \underbrace{X^T X \vec{w}'}_{X^T \vec{y}} + \underbrace{X^T X \vec{n}}_{\vec{0}} = X^T \vec{y}$$

but, $\text{nullsp}(X^T X) = \text{nullsp}(X)$

$$X = \begin{bmatrix} 3 & 1 & 0 \\ 6 & 2 & 1 \\ 3 & 1 & 0 \end{bmatrix}$$

$$\operatorname{rank}(X) = 2$$

$$\dim(\operatorname{nullsp}(X)) = 1$$

$$X \underbrace{\begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}}_{\vec{v}} = \vec{0}$$

$$\operatorname{nullsp}(X) = \operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix} \right\}\right)$$

$$X = \begin{bmatrix} 3 & 1 & 0 \\ 6 & 2 & 1 \\ 3 & 1 & 0 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

$\Rightarrow$ one solution to $X^T X \vec{w}^* = X^T \vec{y}$ is

$$\vec{w}' = \begin{bmatrix} 0 \\ 5/2 \\ -4 \end{bmatrix},$$

$$X' = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\vec{w}' = (X'^T X')^{-1} X'^T \vec{y}$$

so the set of all solutions to $X^T X \vec{w}^* = X^T \vec{y}$ is

not a subsp!

$$\left\{ \begin{bmatrix} 0 \\ 5/2 \\ -4 \end{bmatrix} + t \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}, \ t \in \mathbb{R} \right\}$$

$$A\vec{v} = \vec{0}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$R_{sq}(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right)^2$$

Ch. 3.1

$$= \frac{1}{n}\|\vec{y} - \underbrace{X\vec{w}}_{=\vec{p}}\|^2$$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

"design matrix"

$$X\vec{w} = \begin{bmatrix} w_0 + w_1 x_1 \\ w_0 + w_1 x_2 \\ \vdots \end{bmatrix}$$

Chapter 3.2

3:47

"x_i" "y_i"

departure_hour	day	day_of_month	minutes
10.816667	Mon	15	68.0
7.750000	Tue	16	94.0
8.450000	Mon	22	63.0
7.133333	Tue	23	100.0
9.150000	Tue	30	69.0

Multiple linear regression

$h(\text{departure\_hour}_i, \text{dom}_i) =$

$\vec{w}^*$: solve the normal equations!

$$h(\text{departure hour}_i, \text{dom}_i)$$

$$= w_0 + w_1 \cdot \text{departure hour}_i + w_2 \cdot \text{dom}_i$$

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & dh_1 & dom_1 \\ 1 & dh_2 & dom_2 \\ \vdots & \vdots & \vdots \\ 1 & dh_n & dom_n \end{bmatrix} \quad \vec{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

General

d features

$x_i^{(j)}$ feature j for row i

$h(\vec{x}_i) = w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + \cdots + w_d x_i^{(d)}$

multiple linear regression

d features $\rightarrow$ $d+1$ parameters

departure_hour	day	day_of_month
10.816667	Mon	15
7.750000	Tue	16
8.450000	Mon	22
7.133333	Tue	23
9.150000	Tue	30

$x^{(1)}$ $\quad$ $x^{(2)}$

"feature vector"

$$\vec{x}_i = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(d)} \end{bmatrix}$$

$$\vec{x}_1 = \begin{bmatrix} 10.816\ldots \\ 15 \end{bmatrix}$$

$$\vec{x}_4 = \begin{bmatrix} 7.133\ldots \\ 23 \end{bmatrix}$$

$$\vec{w} = \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}$$

"augmented feature vector"

$$\text{Aug}(\vec{x}_i) = \begin{bmatrix} 1 \\ \vec{x}_i \end{bmatrix}$$

for intercept

$$\text{Aug}(\vec{x}_2) = \begin{bmatrix} 1 \\ 7.75 \\ 16 \end{bmatrix}$$

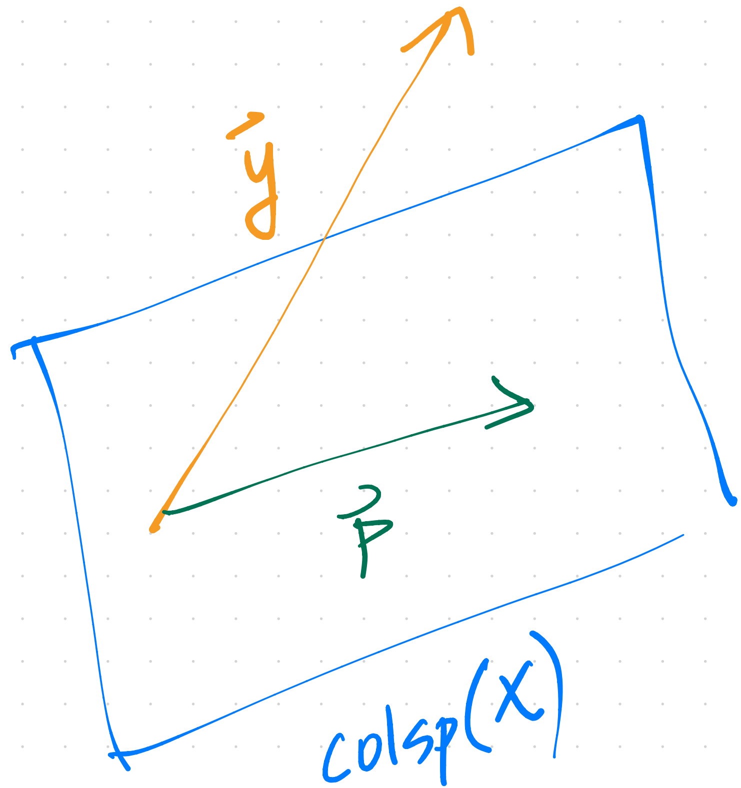
feature vector

$$h(\vec{x}_i) = w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + \dots + w_d x_i^{(d)}$$

$$= \underbrace{\vec{w}}_{\substack{d+1 \\ \text{entries}}} \cdot \underbrace{\text{Aug}(\vec{x}_i)}_{\substack{d+1 \\ \text{entries}}}$$

how do we find $w_0^*, w_1^*, \dots, w_d^*$?

normal equation



$\vec{y}$
$\vec{p}$
colsp(X)

$$X_{n \times (d+1)} = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & & x_1^{(d)} \\ 1 & x_2^{(1)} & x_2^{(2)} & & x_2^{(d)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & & x_n^{(d)} \end{bmatrix}$$

$$R_{sq}(\vec{w}) = \frac{1}{n} \|\vec{y} - X\vec{w}\|^2$$

departure_hour	day	day_of_month	minutes
10.816667	Mon	15	68.0
7.750000	Tue	16	94.0
8.450000	Mon	22	63.0
7.133333	Tue	23	100.0
9.150000	Tue	30	69.0

$$\boxed{f(x) = ax^2 + bx + c}$$

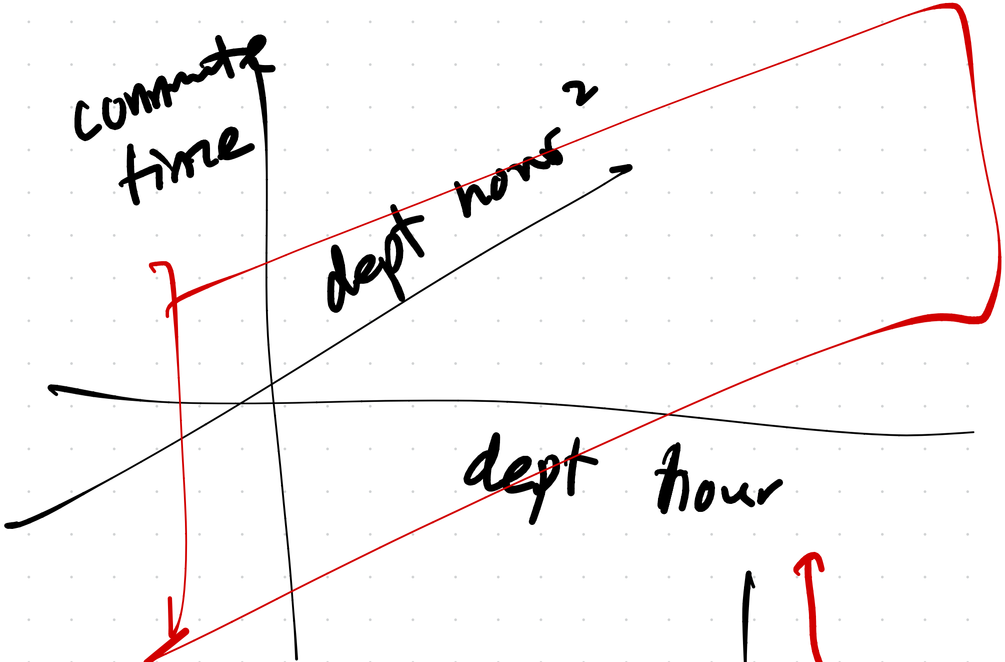
possible

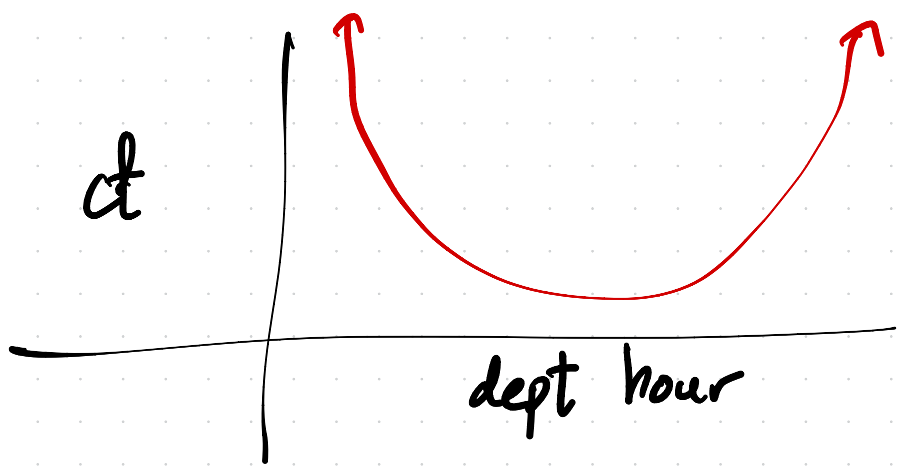
$$g(x) = e^{-cx^2}$$

x : departure time

$$h(x_i) = w_0 + w_1 x_i + w_2 x_i^2 = \vec{w} \cdot \underbrace{\begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix}}_{\text{dot product}}$$

$$X = \begin{bmatrix} 1 & dh_1 & dh_1^2 \\ 1 & dh_2 & dh_2^2 \\ \vdots & \vdots & \vdots \\ 1 & dh_n & dh_n^2 \end{bmatrix}$$


commute time
dept hour²
dept hour


ct
dept hour

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

$$\downarrow$$

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$$

day
Mon
Tue
Mon
...
Mon
Tue
Thu

day == Mon	day == Tue	day == Wed	day == Thu	day == Fri
$\boxed{1}$	0	0	0	0
0	$\boxed{1}$	0	0	0
$\boxed{1}$	0	0	0	0
...	...	...	...	...
$\boxed{1}$	0	0	0	0
0	$\boxed{1}$	0	0	0
0	0	0	$\boxed{1}$	0

"one hot encoding"