EECS 245, Winter 2026

LEC 2

Empirical Risk Minimization and Simple Linear Regression

→ Read Ch. 1.3, 1.4,
all of Ch. 2

Agenda

Keep up with the readings!

- Recap: Modeling recipe
- Comparing loss functions
  - → Squared vs. absolute
  - → Outliers
  - → Other loss functions

Ch. 1.3 + 1.4

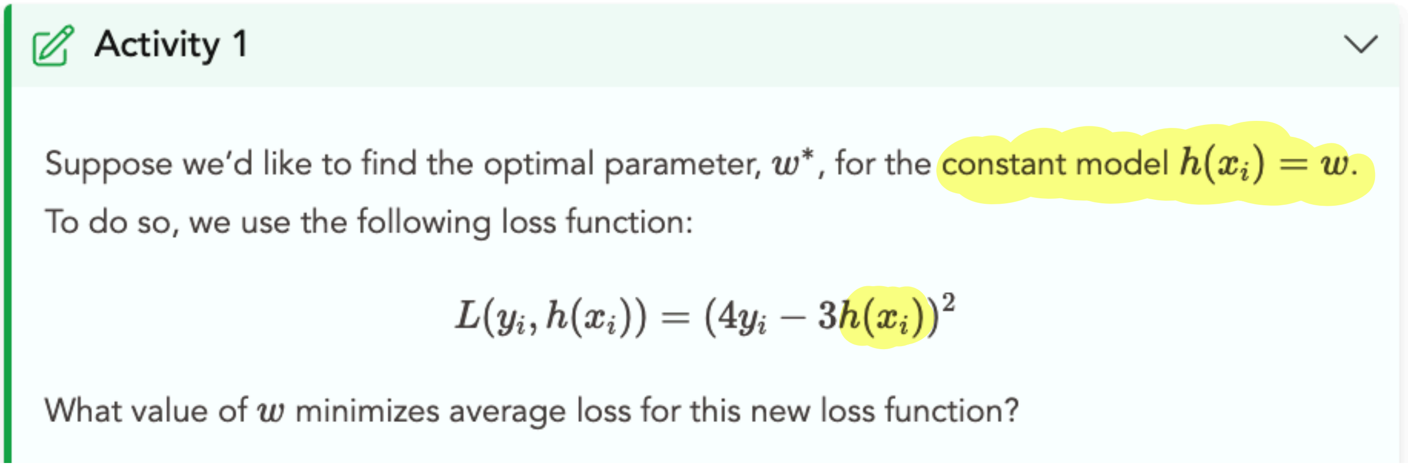
- All about simple linear regression

Ch. 2

Announcements

- HW 1 due Sunday (+survey), HW 2 due Wednesday
- Lab 1 solutions up; new format
- Resources page has past exam Q's sorted by topic
- New Midterm 1 date: Friday, May 22nd, 1-3PM

confirm that it works!



Activity 1

Suppose we'd like to find the optimal parameter, w^*, for the constant model $h(x_i) = w$.

To do so, we use the following loss function:

$$L(y_i, h(x_i)) = (4y_i - 3h(x_i))^2$$

What value of w minimizes average loss for this new loss function?

$$R(w) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}}_{\text{average}} (4y_i - 3w)^2$$

To find the best w^*, we need to minimize this R.

shortcut in this case

$$R(w) = \frac{1}{n} \sum_{i=1}^{n} (4y_i - 3w)^2$$

idea: let $z_i = 4y_i$, and $t = 3w$

mean of $z_1, ---, z_n$

$$R(t) = \frac{1}{n} \sum_{i=1}^{n} (z_i - t)^2 \Rightarrow t^* = \bar{z}$$

$$t^* = 4\bar{y}$$

$$3w^* = 4\bar{y}$$

$$w^* = \frac{4}{3}\bar{y}$$

Three-step modeling recipe "empirical risk minimization"

① Choose a model

constant model, $h(x_i) = \overset{\text{parameter}}{\overset{\downarrow}{w}}$

simple linear regression, $h(x_i) = w_0 + w_1 x_i$

② Choose a loss function

squared loss, $L_{sq}(y_i, h(x_i)) = \underset{(\text{actual} - \text{pred})^2}{(y_i - h(x_i))^2}$

absolute loss, $L_{abs}(y_i, h(x_i)) = |y_i - h(x_i)|$

another term for "average loss" is "empirical risk"

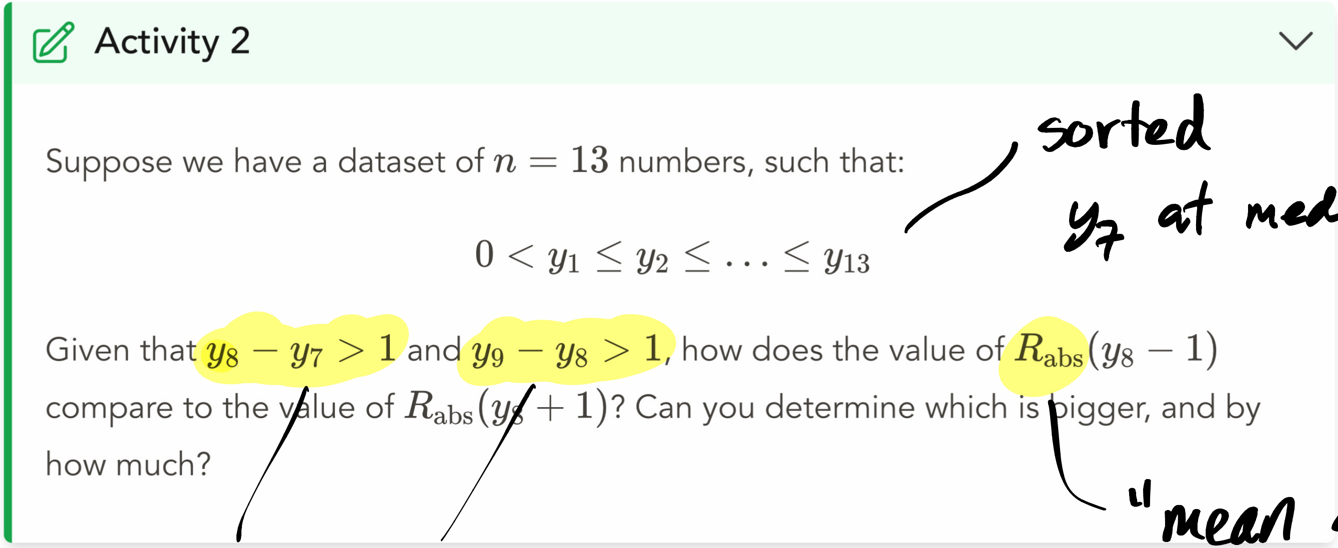
③ Minimize average loss to find optimal parameters

→ If we use squared loss + constant model,

average loss:

$$R_{sq}(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w)^2$$

→ when using squared loss,
average loss = "mean squared error"



Activity 2

Suppose we have a dataset of $n = 13$ numbers, such that:

$$0 < y_1 \leq y_2 \leq \ldots \leq y_{13}$$

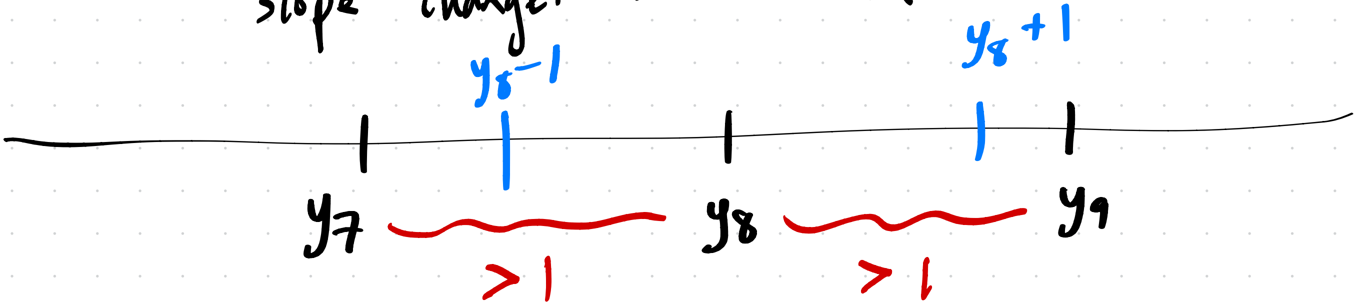
Given that $y_8 - y_7 > 1$ and $y_9 - y_8 > 1$, how does the value of $R_{\text{abs}}(y_8 - 1)$ compare to the value of $R_{\text{abs}}(y_8 + 1)$? Can you determine which is bigger, and by how much?

sorted

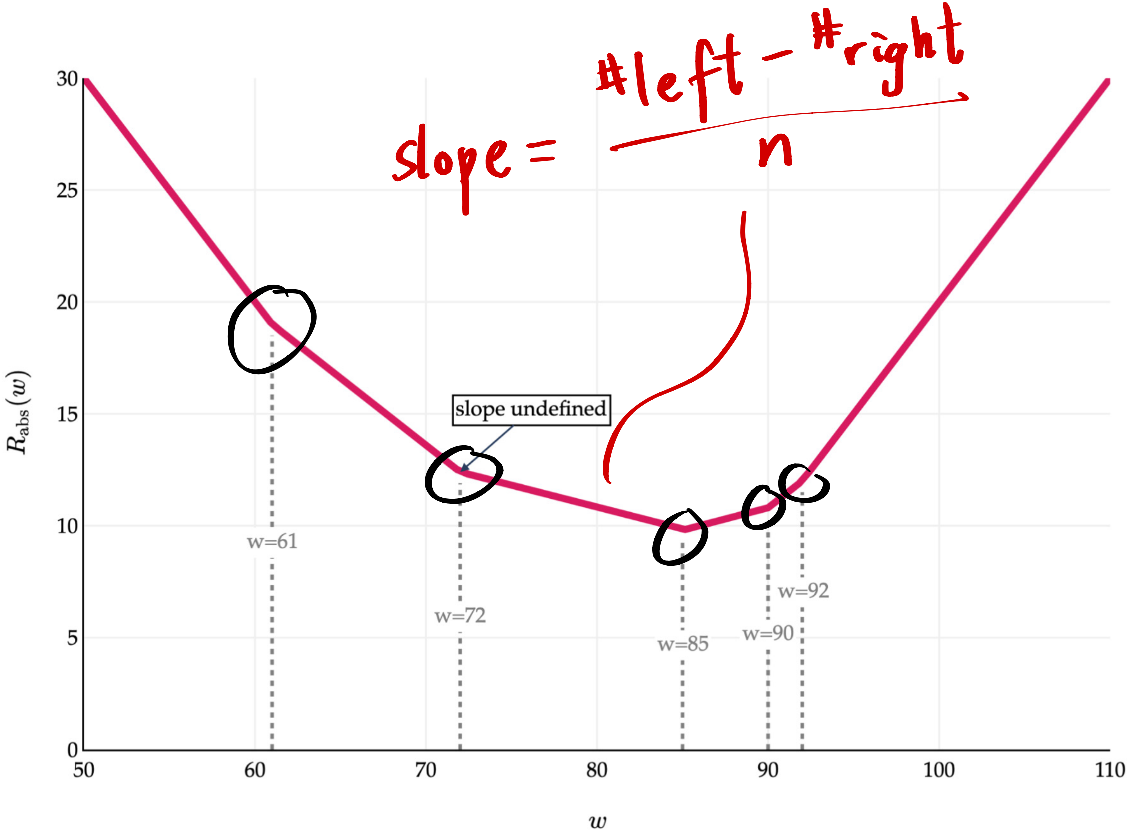
y_7 at median

"mean absolute error"

key: y_7, y_8, y_9 are all distinct; slope changes at each one

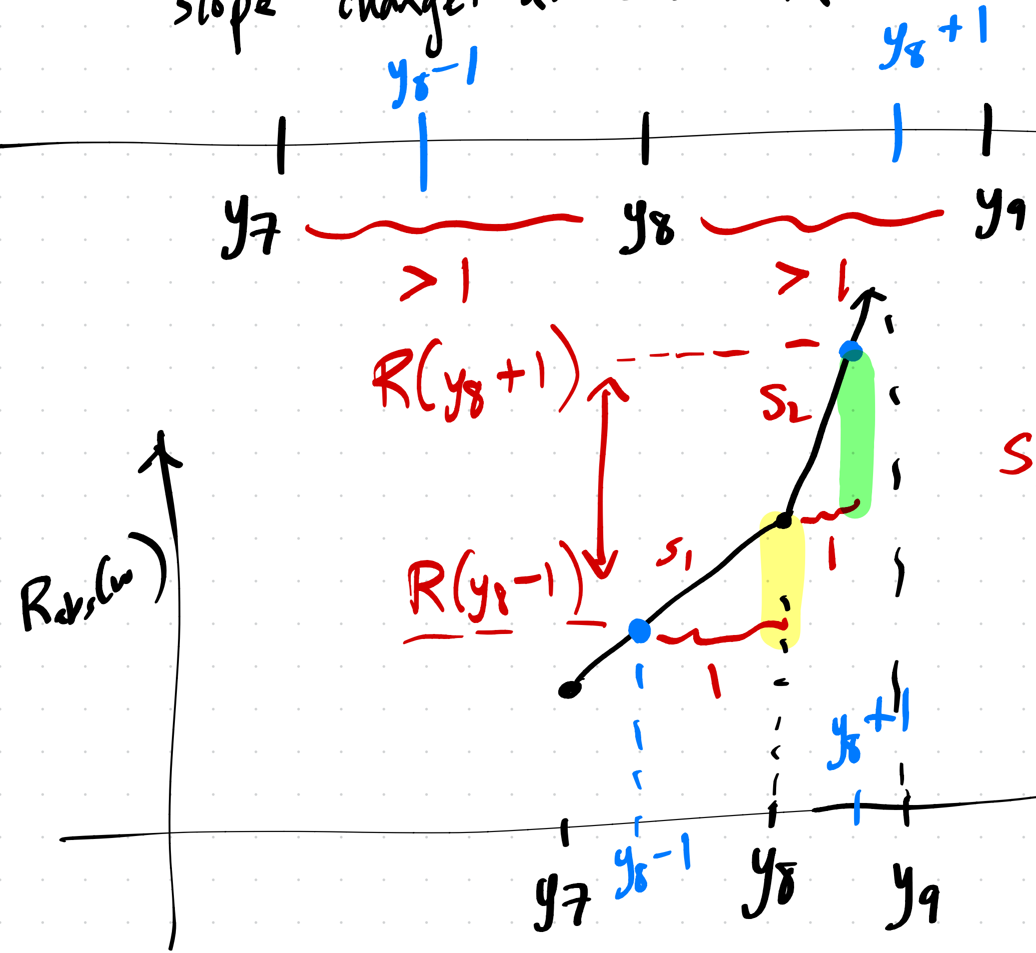


$$R_{\text{abs}}(w) = \frac{1}{5}(|72 - w| + |90 - w| + |61 - w| + |85 - w| + |92 - w|)$$



$$\text{slope} = \frac{\#\text{left} - \#\text{right}}{n}$$

slope changes at each one



Total change:

$$\frac{1}{13}+\frac{3}{13}=\boxed{\frac{4}{13}}$$

so $R(y_8+1)-R(y_8-1)=\frac{4}{13}$

$R(y_8+1)$ is bigger!

S_1 = slope between y_7, y_8 $=\dfrac{7-6}{13}$

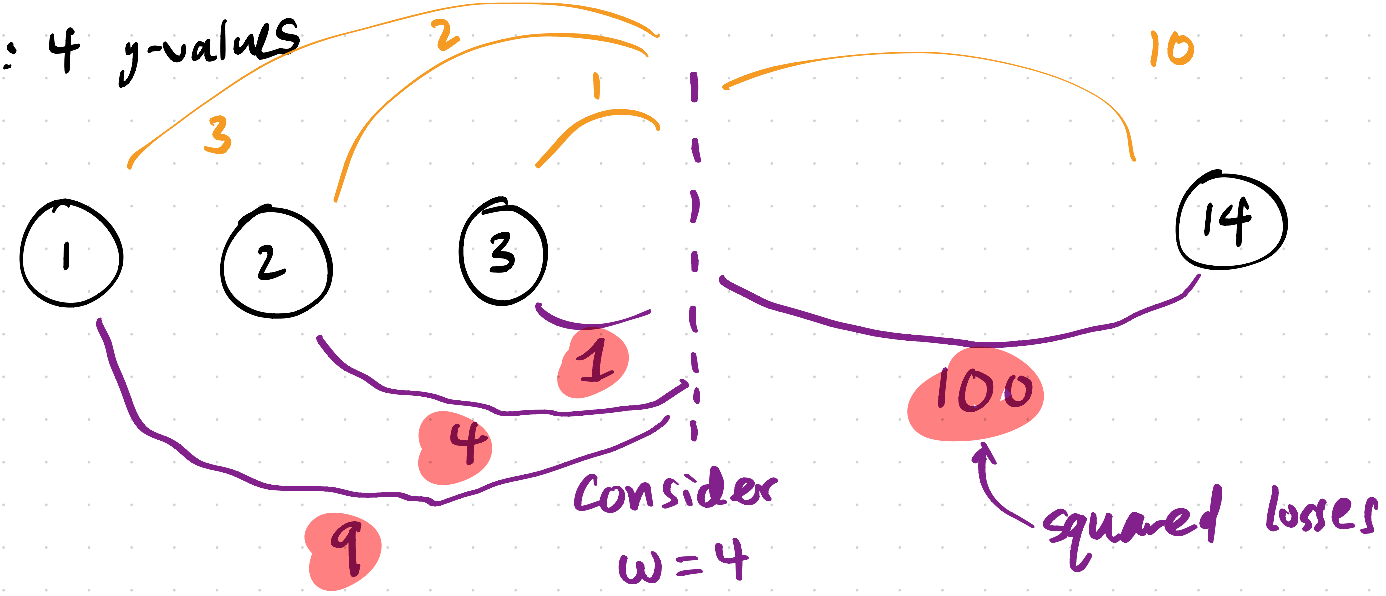
$=\dfrac{1}{13}$

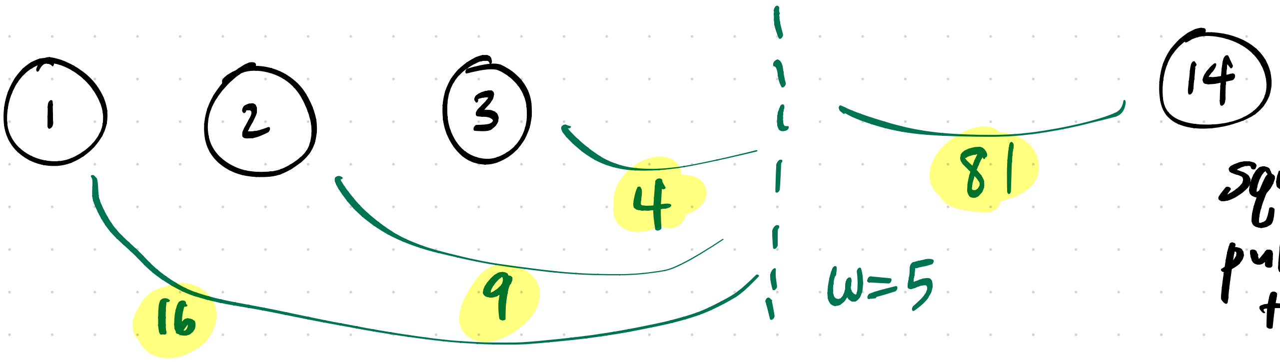
$S_2=\dfrac{8-5}{13}=\dfrac{3}{13}$

Ch. 1.4

- Mean minimizes Mean Squared Error
- Median minimizes Mean Absolute Error
- Mean is more affected by outliers..... why?

example: 4 y-values





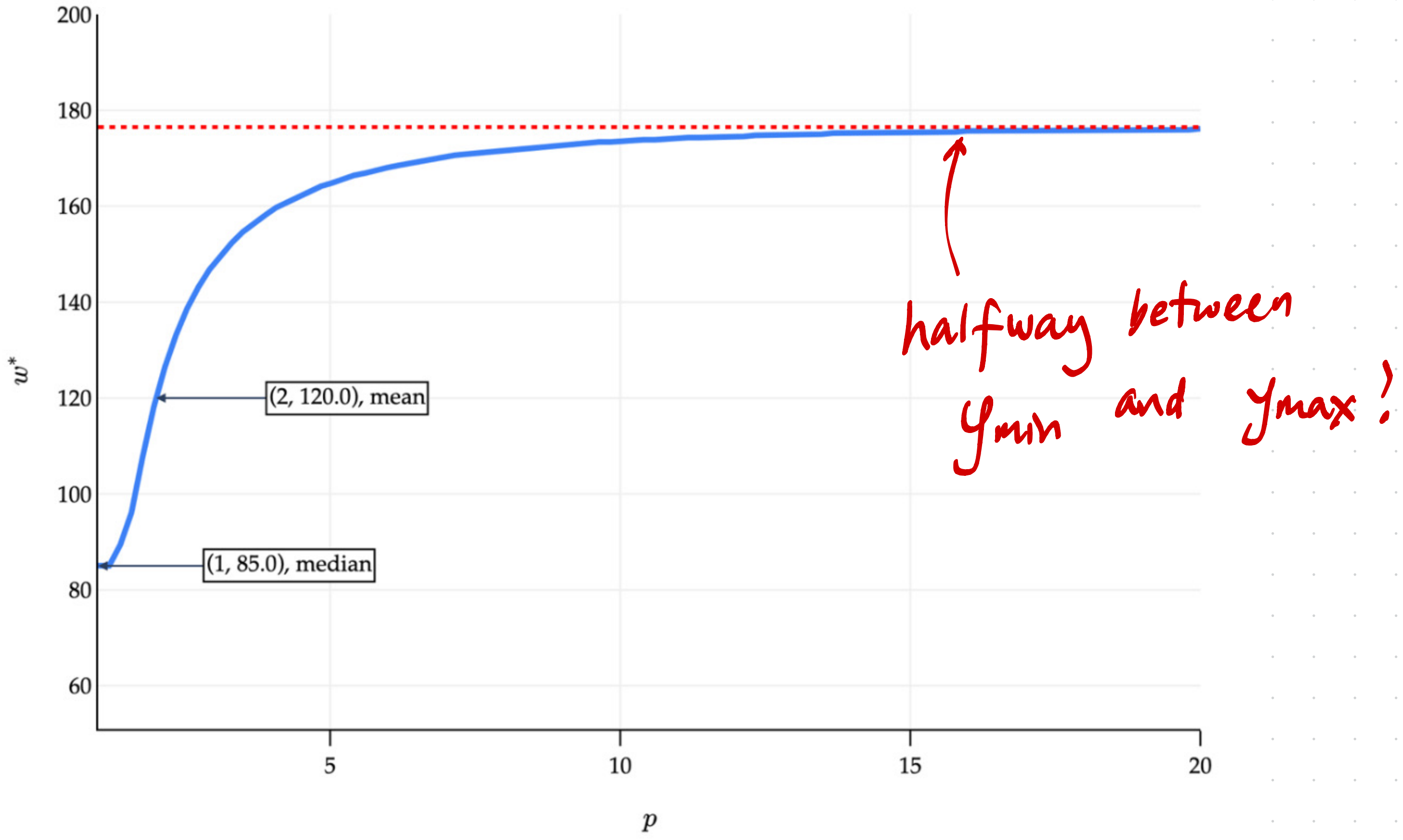
squared loss pulls us towards outliers!

① ② ③ ⑭

what if we use degree-4 loss? or degree-100?

$$R(w) = \lim_{p\to\infty} \frac{1}{n}\sum_{i=1}^{n} |y_i - w|^p$$

$\Rightarrow$ answer: $\frac{y_{min} + y_{max}}{2}$ "midrange"



halfway between y_{min} and y_{max} !

Loss	Minimizer of Empirical Risk	Always Unique?	Robust to Outliers?	Empirical Risk Differentiable?
✓ L_{sq}	mean	yes ✅	no ❌	yes ✅
✓ L_{abs}	median	no ❌	yes ✅	no ❌
✓ L_{∞}	midrange	yes ✅	no ❌	no ❌
$L_{0,1}$	mode	no ❌	no ❌	no ❌

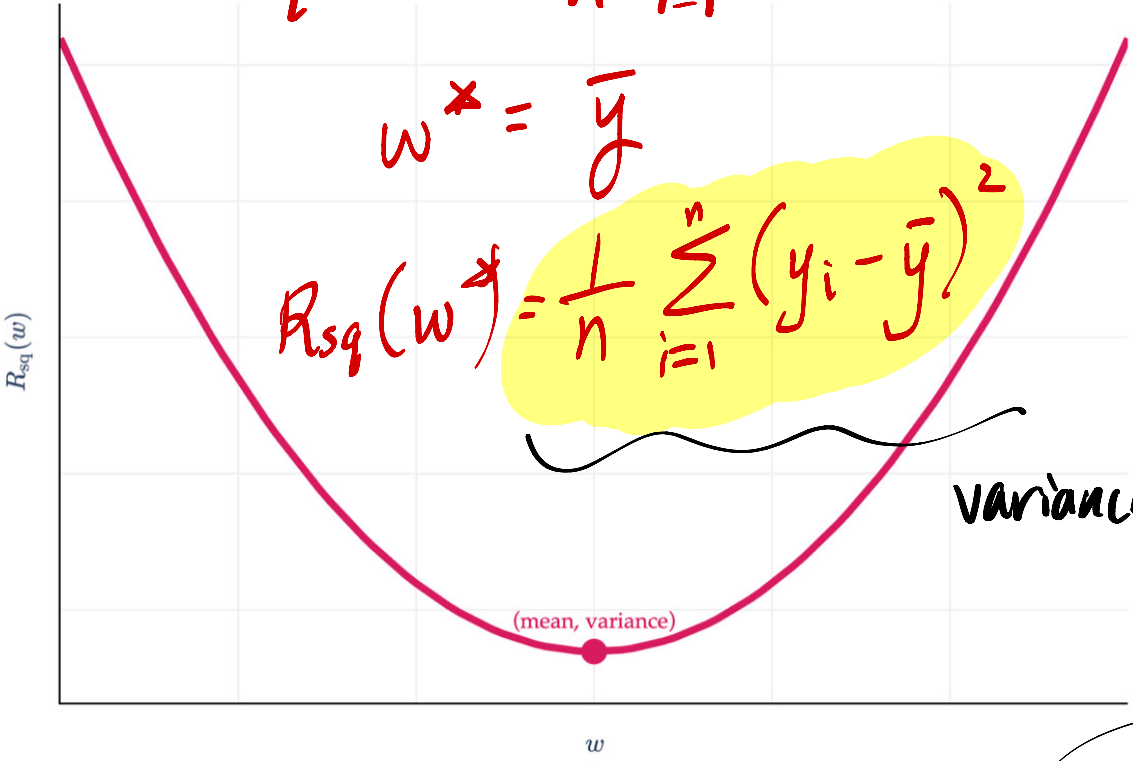
$$L_{0,1}(y_i, w) = \begin{cases} 0 & \text{if } y_i = w \\ 1 & \text{otherwise} \end{cases}$$


Frequency
0
2
4
6
8
10
12
14
Median
Mean
60
70
80
90
100
110
120
130
Commute Time (Minutes)
tail of distribution
right-tailed
right-skewed

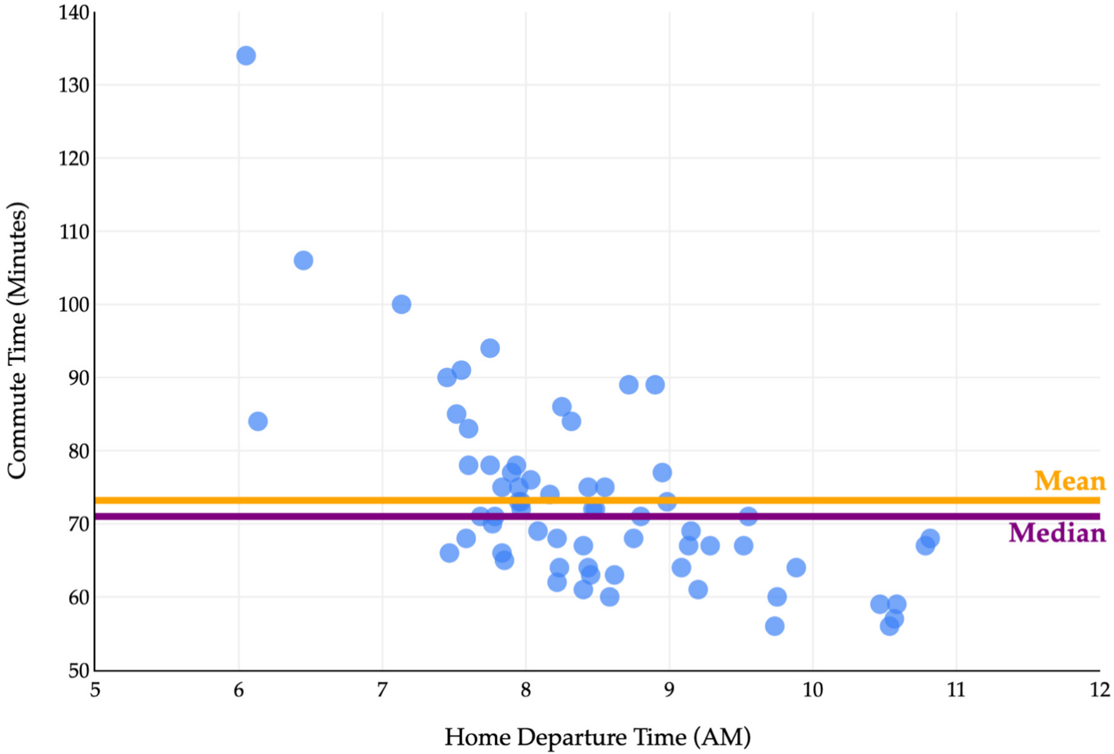
$$R_{sq}(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w)^2$$

$$w^* = \bar{y}$$

$$R_{sq}(w^*) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{variance of } y\text{'s!}}$$

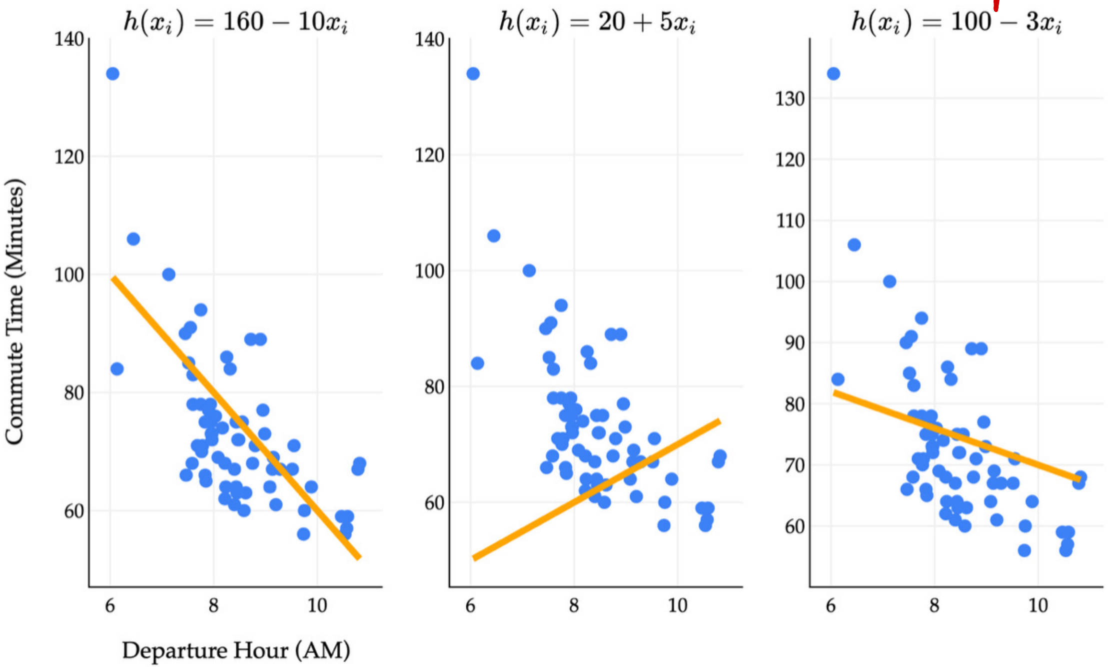


2:22


Commute Time (Minutes)
140
130
120
110
100
90
80
70
60
50
Mean
Median
5
6
7
8
9
10
11
12
Home Departure Time (AM)

Chapter 2: Simple linear regression

$$h(x_i) = \overset{\text{intercept}}{w_0} + \underset{\text{slope}}{w_1} x_i$$



Three-step recipe strikes again!

one input: "simple"
multiple inputs: "multiple"

① Choose a model: $h(x_i) = w_0 + w_1 x_i$

simple linear regression

② Choose a loss function:

$$L_{sq}(y_i, h(x_i)) = (y_i - h(x_i))^2$$

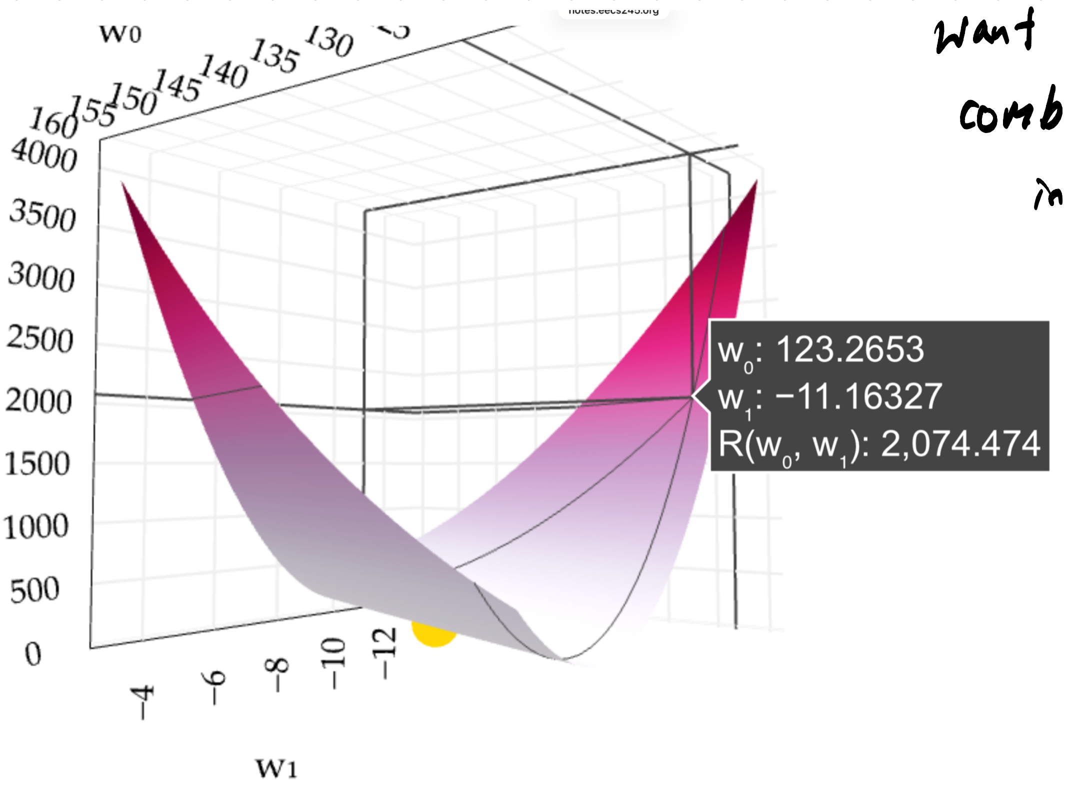
③ Minimize average loss to find optimal parameters

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

Two parameters : w_0 and w_1

What does R_{sq} even look like?

$\Rightarrow$ looks like a bowl in 3D!



want to find the combination of intercept (w_0) and slope (w_1) at the bottom of the loss surface

Chapter 2.2: Detour: Partial Derivatives

e.g. $$f(x, y) = \frac{x^2+y^2}{9}$$

Suppose f is a function of multiple input variables, e.g. x, y, z.

The partial derivative of f, with respect to y, is taken by treating all other variables as constants.

$$\frac{\partial f}{\partial y}(x, y, z)$$

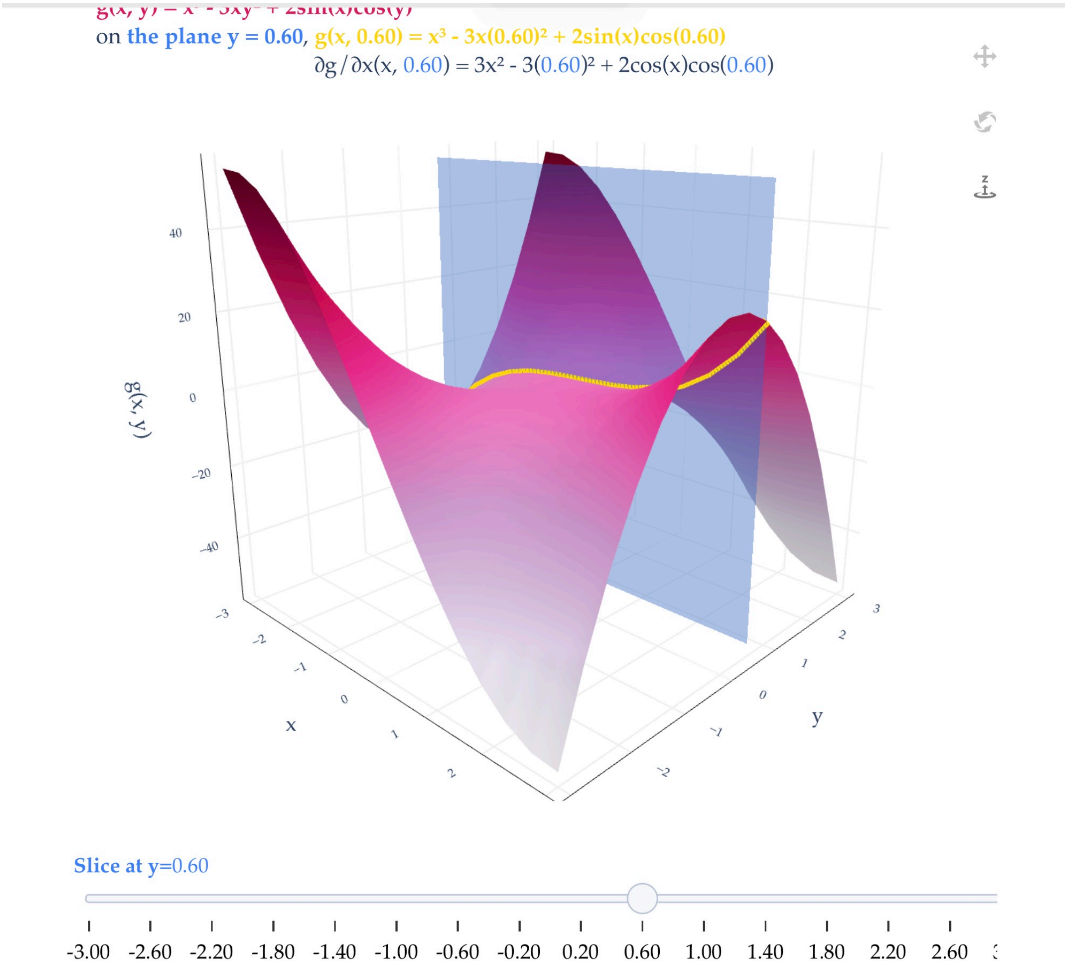
$$f(x,y) = \frac{x^2+y^2}{9} = \frac{1}{9}x^2 + \frac{1}{9}y^2$$

$$\frac{\partial f}{\partial x}(x,y) = \frac{2x}{9} \qquad \frac{\partial f}{\partial y}(x,y) = \frac{2y}{9}$$

Pick a point, e.g. $x=-3$ and $y=1/2$

$$\frac{\partial f}{\partial x}(-3,1/2) = -\frac{6}{9} = -\frac{2}{3} \qquad \frac{\partial f}{\partial y}(-3,1/2) = 1/9$$

$$\approx -0.67 \qquad \approx 0.11$$



$g(x, y) = x^3 - 3xy^2 + 2\sin(x)\cos(y)$

on **the plane y = 0.60**, $g(x, 0.60) = x^3 - 3x(0.60)^2 + 2\sin(x)\cos(0.60)$

$\partial g / \partial x(x, 0.60) = 3x^2 - 3(0.60)^2 + 2\cos(x)\cos(0.60)$

Slice at y=0.60

Go play with slider in Ch. 2.2!

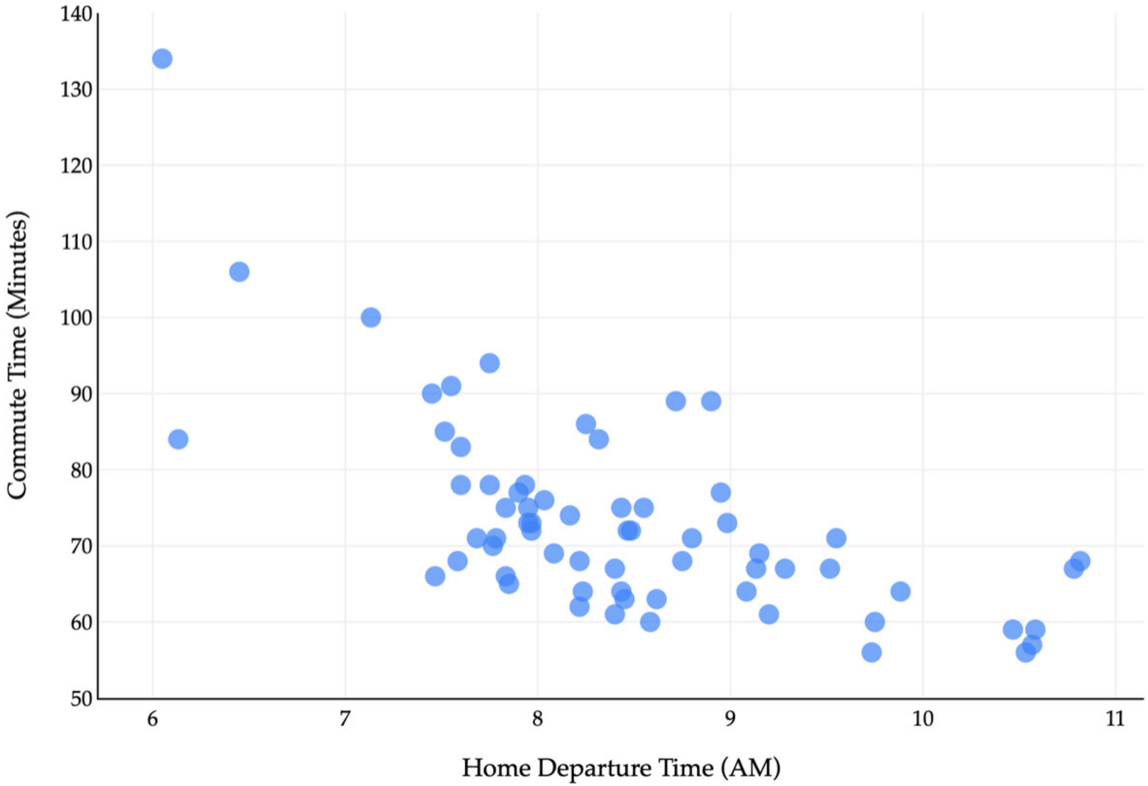
How do we minimize a function with multiple input variables?

⇒ answer: set all partial derivatives equal to 0, solve resulting system of equations!

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

Plan: ① Compute $\frac{\partial R}{\partial w_0}$ and $\frac{\partial R}{\partial w_1}$

② Set both equal to 0 and solve for the resulting w_0^*, w_1^*


140
130
120
110
100
90
80
70
60
50
Commute Time (Minutes)
6
7
8
9
10
11
Home Departure Time (AM)

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

$$\frac{\partial R}{\partial w_0} = \frac{1}{n}\sum_{i=1}^{n}\underline{2}\left(y_i - (w_0 + w_1 x_i)\right)\underline{(-1)}$$

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

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

$$\frac{\partial R}{\partial w_0} = -\frac{2}{n}\sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right) = 0$$

$$\frac{\partial R}{\partial w_1} = -\frac{2}{n}\sum_{i=1}^{n}x_i\left(y_i - (w_0 + w_1 x_i)\right) = 0$$

$$\sum_{i=1}^{n}\left(y_i - (w_0 + w_1 x_i)\right) = 0 \rightsquigarrow$$

English interpretation:

errors add up

to 0!

$$\sum_{i=1}^{n}x_i\left(y_i - (w_0 + w_1 x_i)\right) = 0$$

$$\sum_{i=1}^{n} (y_i - w) = 0 \implies w^* = \bar{y}$$

$$-\frac{2}{n} \sum (y_i - w) \rightarrow \frac{dR_{sq}}{dw} \text{ when}$$

$$R_{sq}(w) = \frac{1}{n} \sum_{i=1}^{n} (y_i - w)^2$$

① $$\sum_{i=1}^{n}\left(y_i-(w_0+w_1x_i)\right)=0$$

② $$\sum_{i=1}^{n}x_i\left(y_i-(w_0+w_1x_i)\right)=0$$

2 equations,
2 unknowns

solve equation ① for w_0, call it w_0^*
then, plug this into ② and solve for w_1^*

$$\sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_i)\right) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} w_0 - \sum_{i=1}^{n} w_1 x_i = 0$$

$$\sum_{i=1}^{n} y_i - n w_0 - w_1 \sum_{i=1}^{n} x_i = 0$$

best intercept depends on the best slope!

$$\underbrace{\frac{\sum_{i=1}^{n} y_i}{n}}_{\bar{y}} - w_1 \underbrace{\frac{\sum_{i=1}^{n} x_i}{n}}_{\bar{x}} = w_0$$

$$\Rightarrow w_0^{*} = \bar{y} - w_1^{*} \bar{x}$$

Idea: plug $w_0^* = \bar{y} - w_1^* \bar{x}$ into

equation ②: $\sum_{i=1}^{n} x_i \left(y_i - (w_0 + w_1 x_i)\right) = 0$

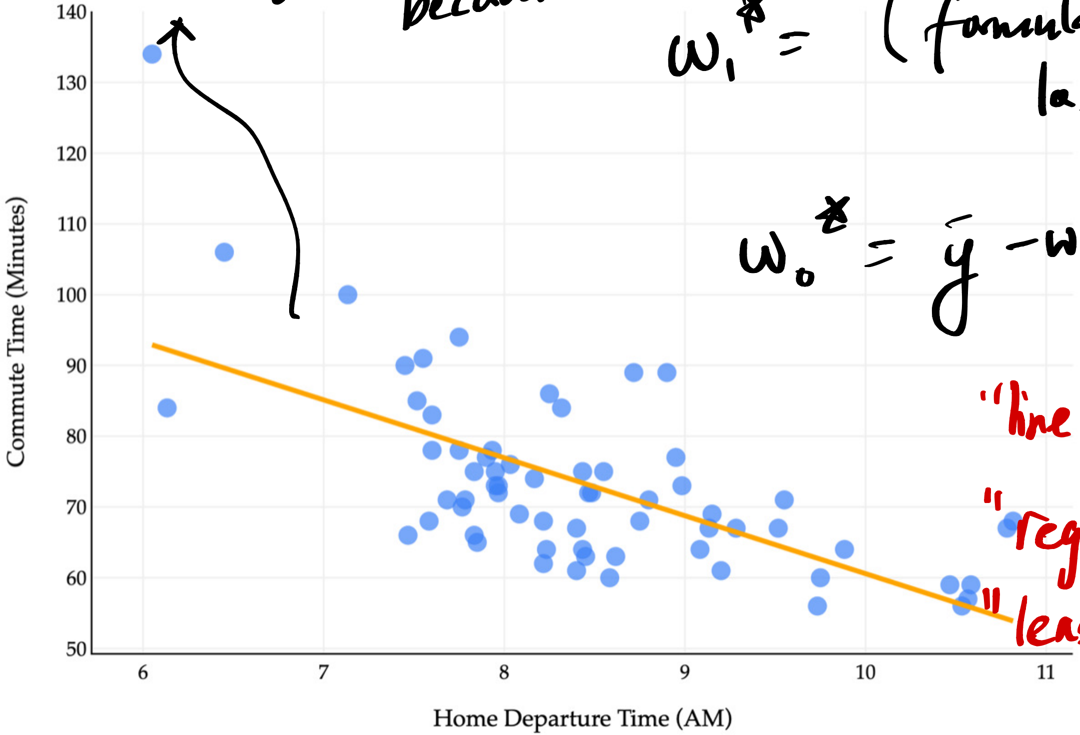
the only unknown then
is w_1^*

$$w_1^* = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})}$$

This line minimizes mean squared error, because its slope is

$w_1^* =$ (formula from last slide),

$w_0^* = \bar{y} - w_1^* \bar{x}$



"line of best fit"

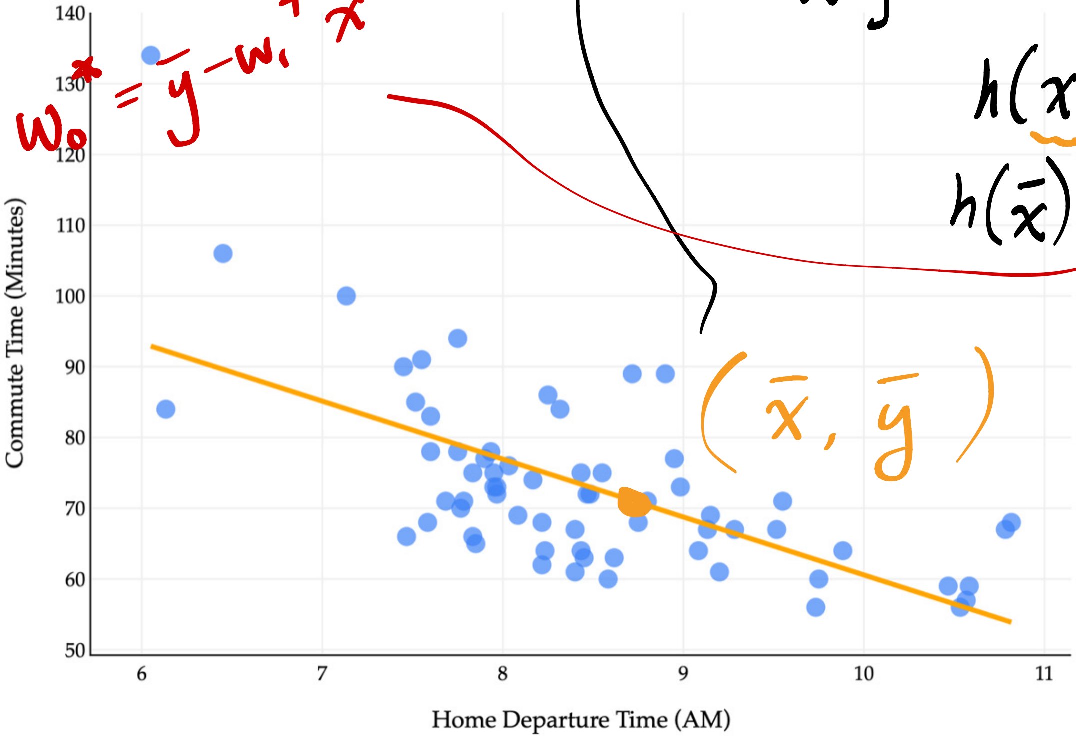
"regression line"

"least squares line"

line guaranteed to pass through $(\bar{x}, \bar{y})$

Why?

$$w_0^* = \bar{y} - w_1^* \bar{x}$$



$$h(x_i) = w_0^* + w_1^* x_i$$

$$h(\bar{x}) = w_0^* + w_1^* \bar{x}$$

$$= \bar{y} \underbrace{- w_1^* \bar{x} + w_1^* \bar{x}}_{=0}$$

$$= \bar{y}$$

The many formulas for the best slope

$$w_1^* = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \boxed{\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

$$= \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

why are all these the same!?

Key fact:

$$\sum_{i=1}^{n}(x_i-\bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}$$

$$= \sum_{i=1}^{n} x_i - n\bar{x}$$

$$= n\bar{x} - n\bar{x}$$

$$= 0 \quad ✓$$

$$\bar{x} = \frac{\sum x_i}{n}$$

$$n\bar{x} = \sum x_i$$

$$\sum_{i=1}^{n}(y_i-\bar{y}) = 0$$

$$\sum_{i=1}^{n} \underbrace{(x_i - \bar{x})}_{\text{distribute}} (y_i - \bar{y})$$

$$= \sum_{i=1}^{n} \Big(x_i (y_i - \bar{y}) - \bar{x} (y_i - \bar{y}) \Big)$$

$$= \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \bar{x} \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})}_{= 0}$$

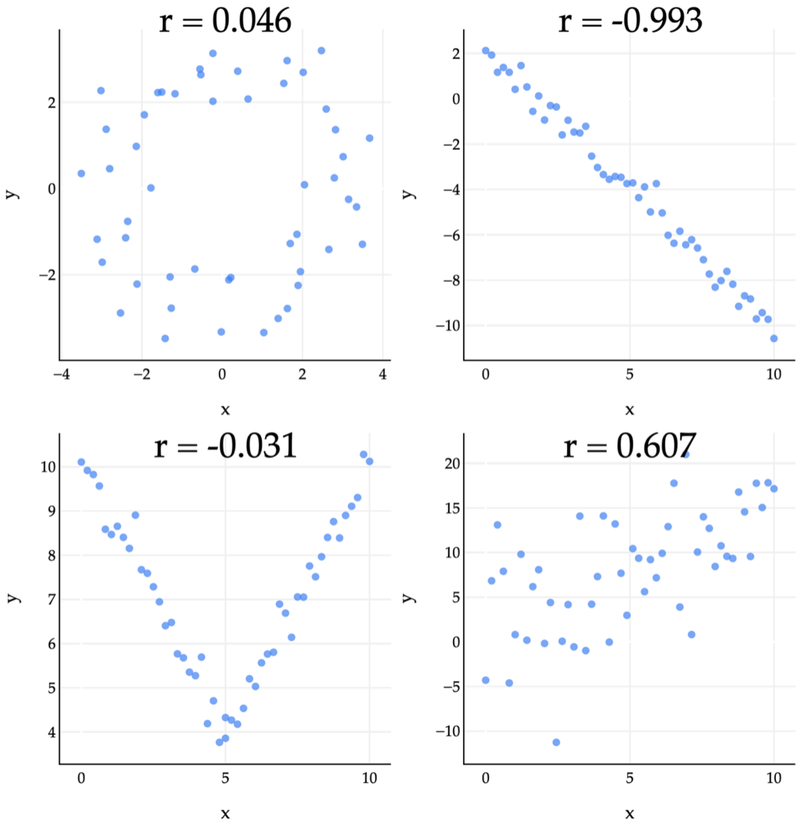
$$= \sum_{i=1}^{n} x_i (y_i - \bar{y})$$

Correlation coefficient

r: measures the strength of the linear association between x and y

$-1 \le r \le 1$

need exact straight line



Another equivalent formula for best slope!

$$w_1^* = r \frac{\sigma_y}{\sigma_x}$$

standard dev of y

$$= \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

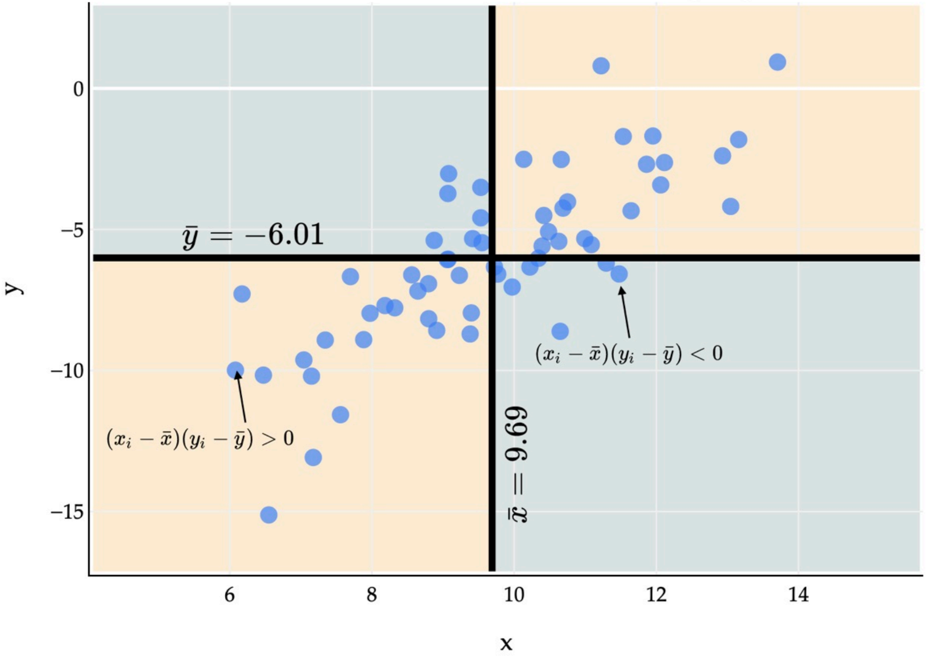
What is the def'n of r?

$$r = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\left(\frac{x_i - \bar{x}}{\sigma_x}\right)}_{\text{\# of SDs above mean } \bar{x}} \left(\frac{y_i - \bar{y}}{\sigma_y}\right)$$

"z-score"

$$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right)$$

= average product of (x_i's z-score) and (y_i's z-score)



r = 0.79

most values are in bottom-left or top-right