EECS 245, Winter 2026

LEC 3

Empirical Risk Minimization; Simple Linear Regression

→ Read: Ch. 1.3, 1.4, 2.1

Agenda

Keep up with the readings!

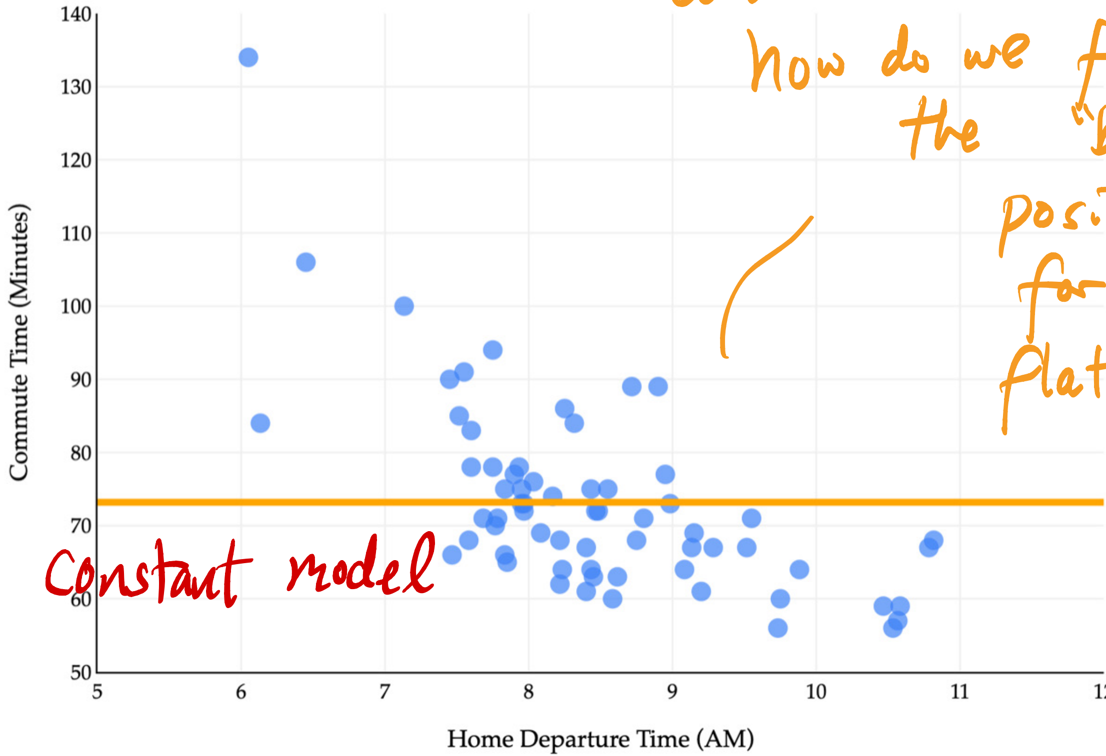
- Recap: The modeling recipe
- "Empirical risk"
- Absolute loss vs. squared loss
  - Minimizing mean absolute error
  - Outliers
  - 0-1 loss? L_∞ loss?
- Intro to simple linear regression

1.3 · 1.4 · 2.1

Announcements

- HW 1 due on Friday, including Welcome Survey
- Office hours after lecture today are in 4174 LEIN
- Lab 2 tomorrow/Friday: don't forget that attendance is required!

Last class: how do we find the "best" position for this flat line?



Three-step modeling recipe

① Choose a model

constant model $h(x_i) = w$

② Choose a loss function

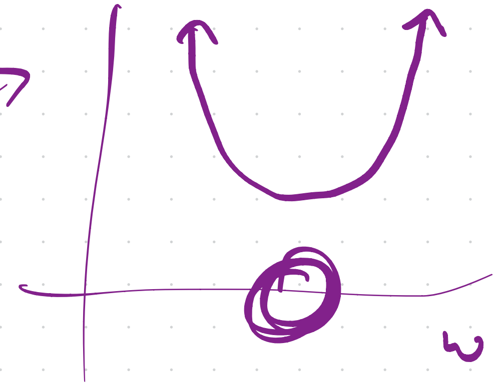
$$L_{sq}(y_i, h(x_i)) = \underbrace{(y_i - h(x_i))^2}_{(\text{actual} - \text{pred})^2}$$

"squared loss"

③ Minimize average loss to find optimal parameters

③ Average loss when we use

- constant model
- squared loss



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

$h(x_i) = w$

function of w only!

$$\Downarrow$$

optimal parameter

$$w^* = \text{Mean}(y_1, y_2, \ldots, y_n) = \bar{y}$$

"Empirical risk" : fancy term for average loss

comes from real data

R_{sq}

"three-step modeling recipe" = "empirical risk minimization"

same meaning

average squared loss = mean squared error

= empirical risk, for squared loss

all mean same thing

Activity

$$L_{Dogs}(y_i, h(x_i)) = (4y_i - 3h(x_i))^2$$

Q: For the constant model, $h(x_i) = w$, find the w^* that minimizes average Dogs loss.

$\underbrace{\text{average Dogs loss}}_{\text{empirical risk}}$

$$R_{Dogs}(w) = \frac{1}{n}\sum_{i=1}^{n}\left(4y_i - 3w\right)^2 \qquad h(x_i) = w$$

Take derivative, set $= 0$, solve for w^*

$$\frac{dR}{dw} = \frac{1}{n}\sum_{i=1}^{n} 2(4y_i - 3w)(-3) = \frac{-6}{n}\sum_{i=1}^{n}(4y_i - 3w)$$

$$\frac{-6}{n}\sum_{i=1}^{n}(4y_i - 3w) = 0$$

$$\sum_{i=1}^{n}(4y_i - 3w) = 0$$

$$4\sum_{i=1}^{n} y_i - 3\sum w = 0$$

$$4\sum_{i=1}^{n} y_i - 3nw = 0 \Rightarrow w^* = \frac{4\sum_{i=1}^{n} y_i}{3n} = \frac{4}{3}\bar{y}$$

substitution

$$z_i = 4y_i$$

$$t = 3w$$

shortcut

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

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

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

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

Absolute loss

Recipe

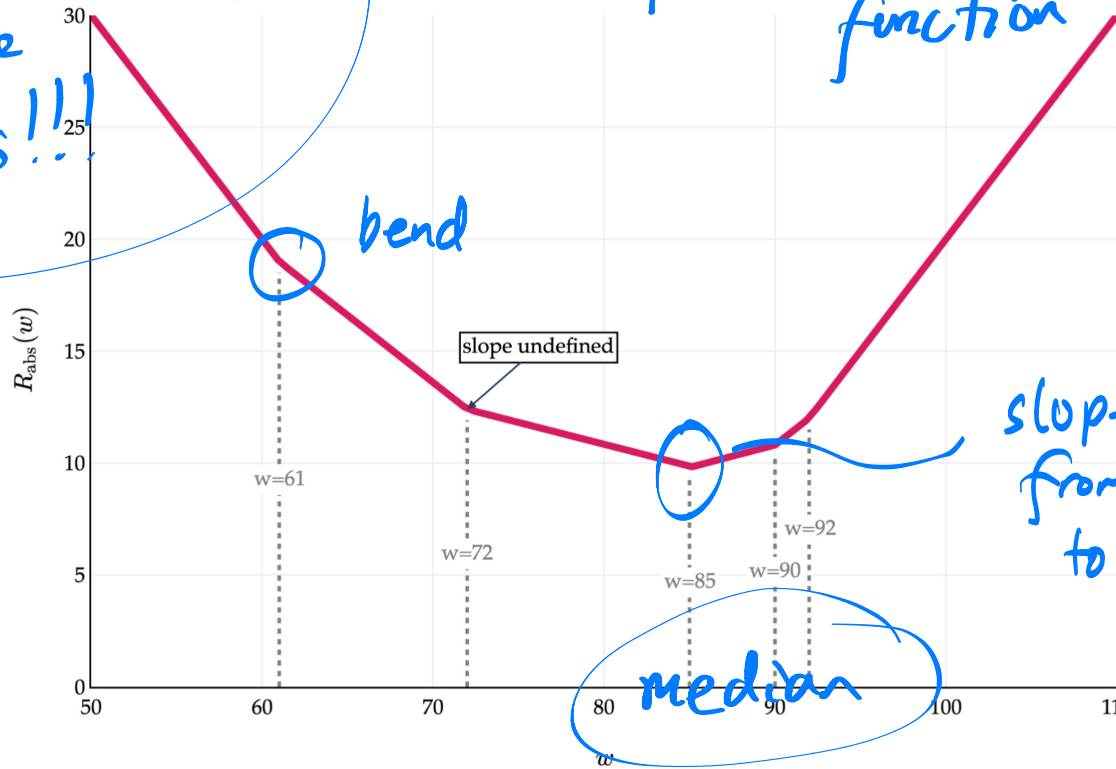
1. Constant model: $h(x_i) = w$
2. Absolute loss: $L_{abs}(y_i, h(x_i)) = |y_i - h(x_i)|$
3. "mean absolute error"

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

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

at each "bend", the slope increases!!!

piecewise linear function



slope goes from negative to positive!

$$R_{\text{abs}}(w) = \frac{1}{6}(|72 - w| + |90 - w| + |61 - w| + |85 - w| + |92 - w| + |78 - w|)$$

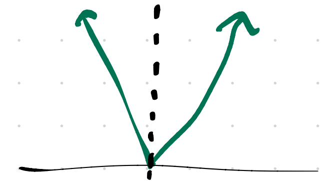
$y_6 = 78$

And its graph is:

"mean absolute error"

Goal: Minimize

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



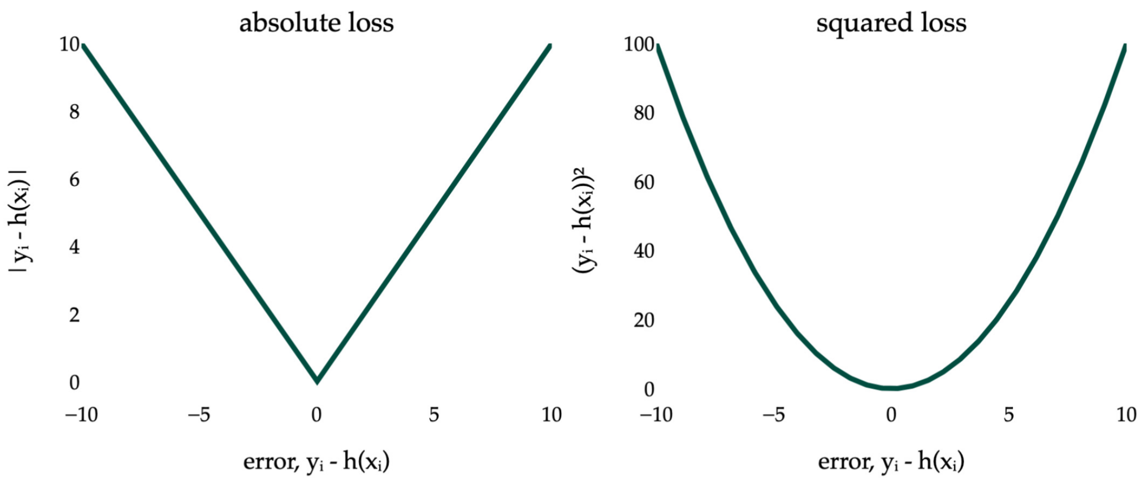
The derivative isn't always defined!

slope

For details, read Ch. 1.3, but the result:

$$\frac{dR_{abs}(w)}{dw} = \frac{(\#\text{ left of } w) - (\#\text{ right of } w)}{n}$$

unless w is a data point, in which case slope not defined!

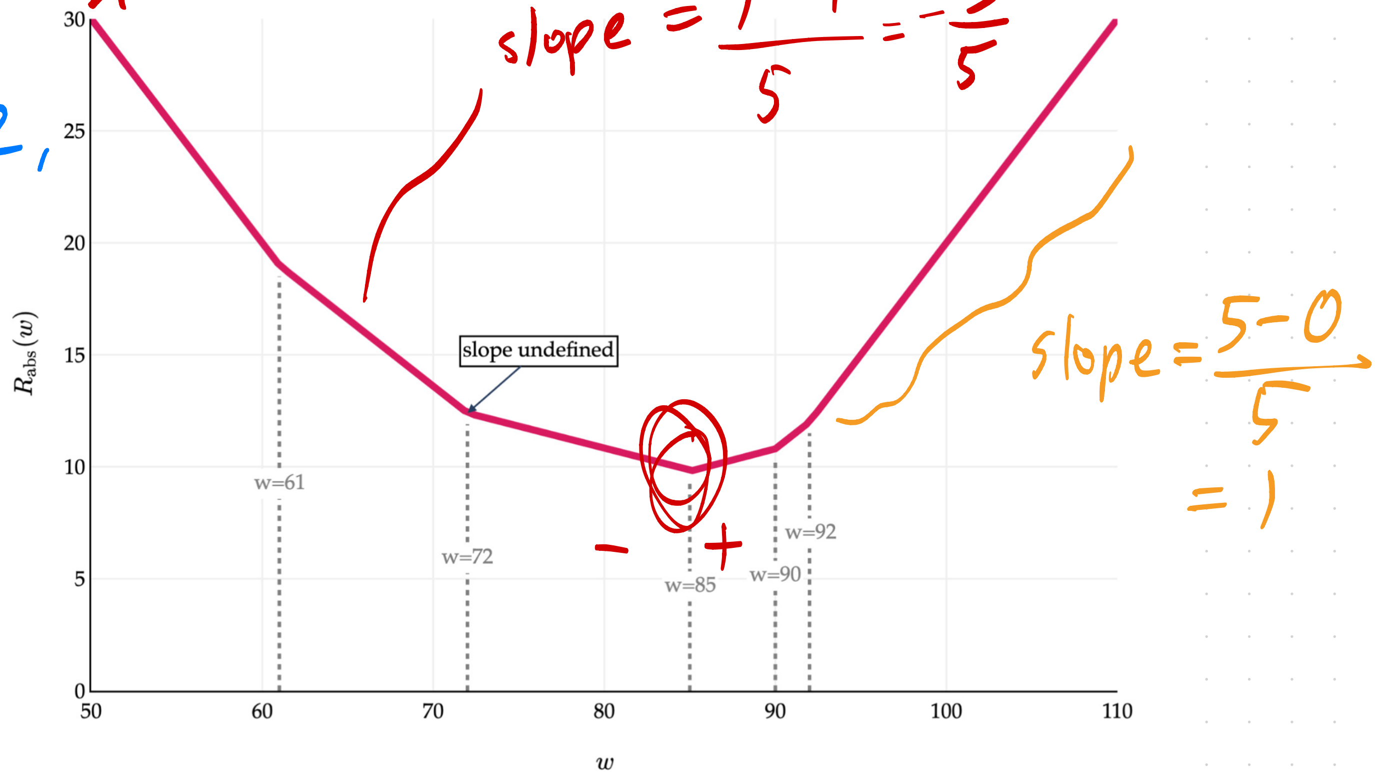


absolute loss
10
8
6
4
2
0
| yi - h(xi) |
−10
−5
0
5
10
error, yi - h(xi)
squared loss
100
80
60
40
20
0
(yi - h(xi))²
−10
−5
0
5
10
error, yi - h(xi)

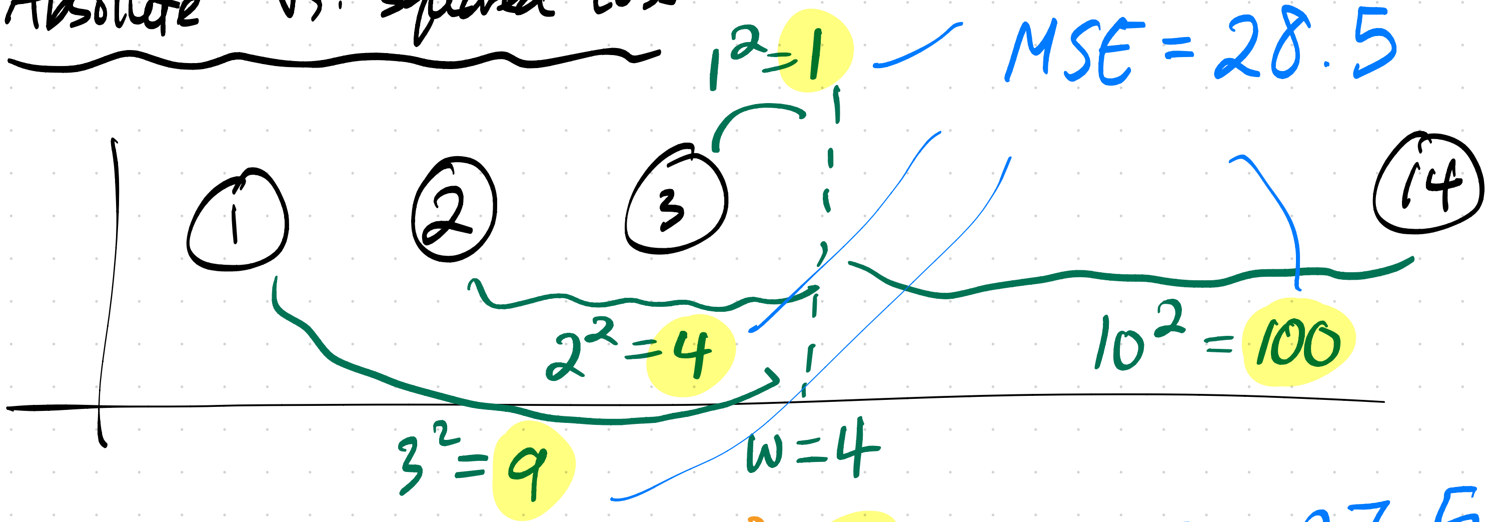
$$R_{\text{abs}}(w) = \frac{1}{5}(|72 - w| + |90 - w| + |61 - w| + |85 - w| + |92 - w|)$$

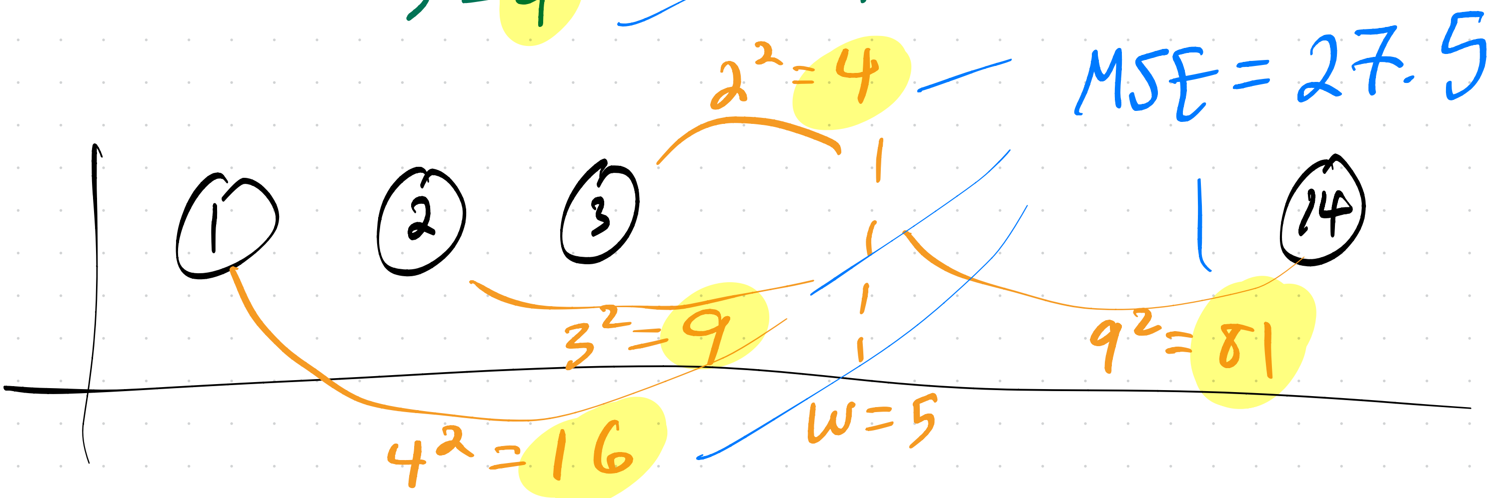
slope $= \frac{\text{left} - \text{right}}{n}$

Activity 2, Ch 1.3



Absolute vs. squared loss





"Balance conditions"

① Median minimizes $R_{abs}(w) = \frac{1}{n}\sum_{i=1}^{n}|y_i - w|$

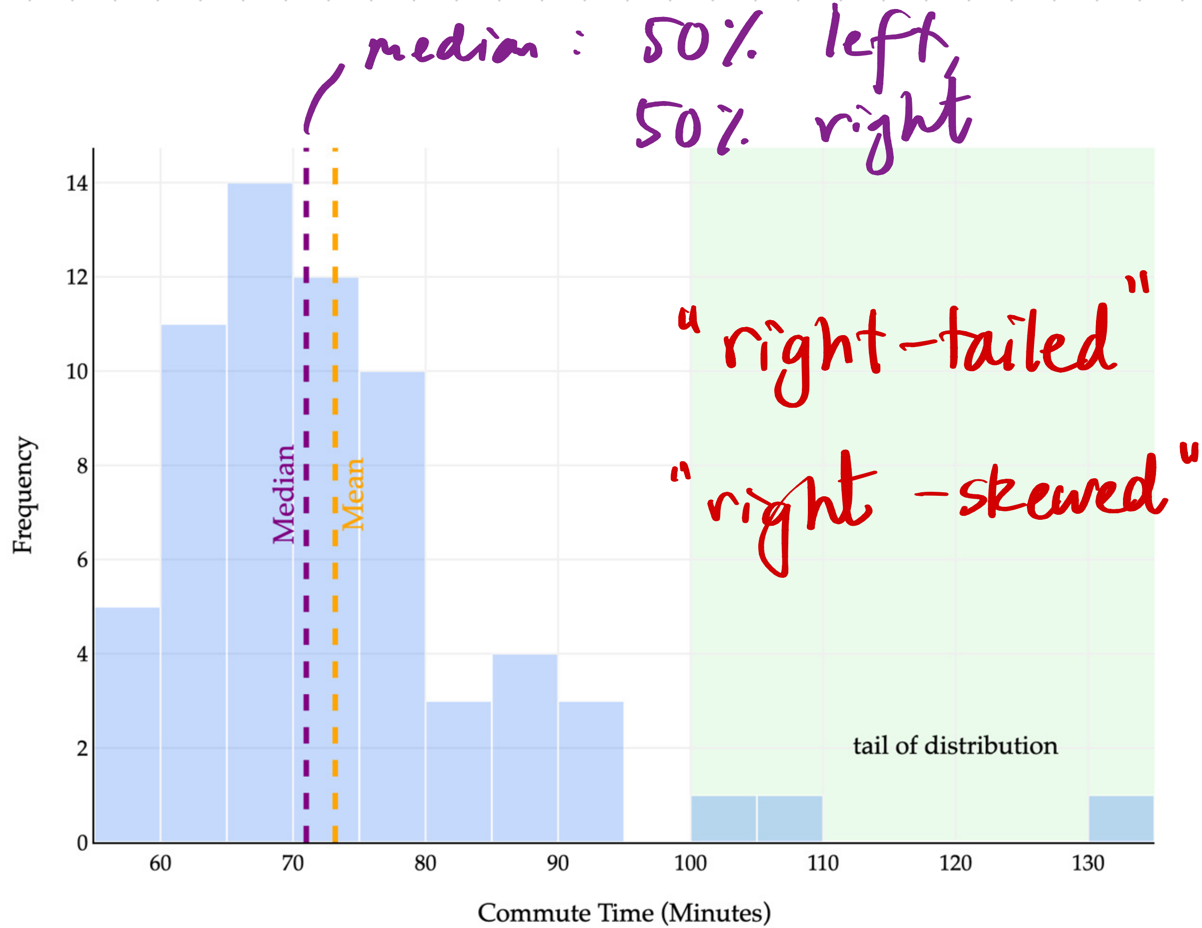
$$\Rightarrow \frac{dR_{abs}}{dw} = \frac{\#\text{left} - \#\text{right}}{n}$$

$\Rightarrow$ median : $\#\text{left} = \#\text{right}$

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

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

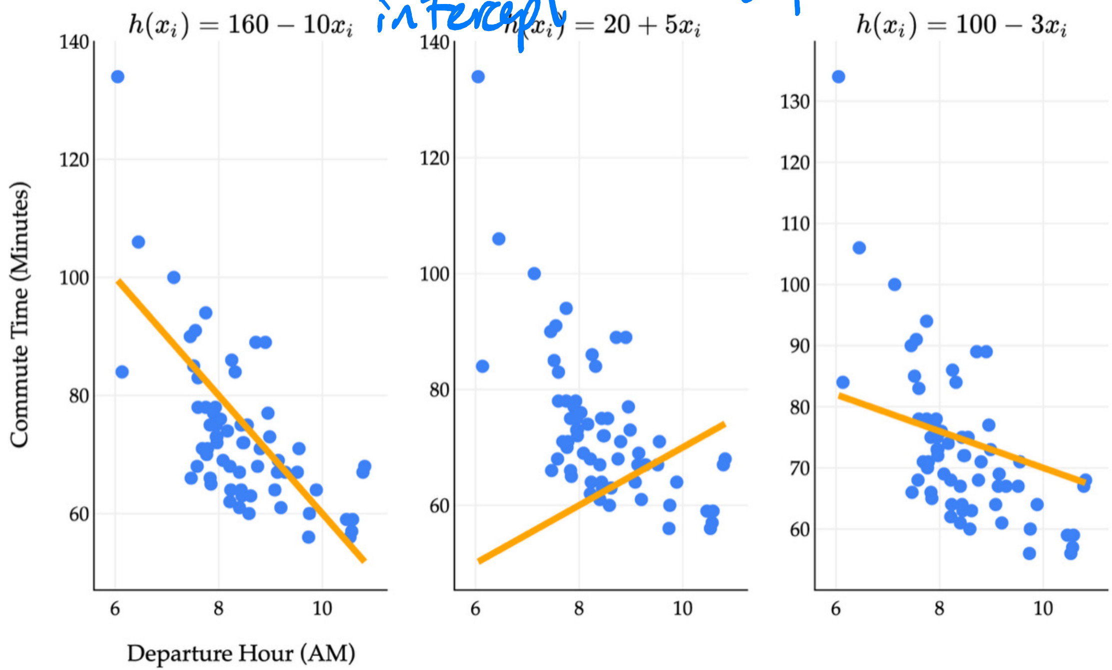
$\sum_{i=1}^{n}(y_i - w) = 0$


median : 50% left, 50% right
"right-tailed"
"right-skewed"
Median
Mean
Frequency
tail of distribution
Commute Time (Minutes)
0
2
4
6
8
10
12
14
60
70
80
90
100
110
120
130

2.1 Simple linear regression

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

w_0 ← intercept, w_1 ← slope



Three-step recipe

1. Choose a model $\qquad h(x_i) = w_0 + w_1 x_i$
2. Choose a loss f'n $\qquad L_{sq}(y_i, h(x_i)) = (y_i - h(x_i))^2$
3. Minimize average loss

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

$(\text{actual} - \text{pred})^2$

Preview:

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

best slope

$$= r \frac{\sigma_y}{\sigma_x}$$

σ_y ← SD of y

r ← "correlation coefficient", $-1 \le r \le 1$

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

best intercept