EECS 245 Fall 2025
Math for ML

Lecture 3: Empirical Risk; Simple Linear Regression

→ Read Ch. 1.3 (new content added)

Agenda

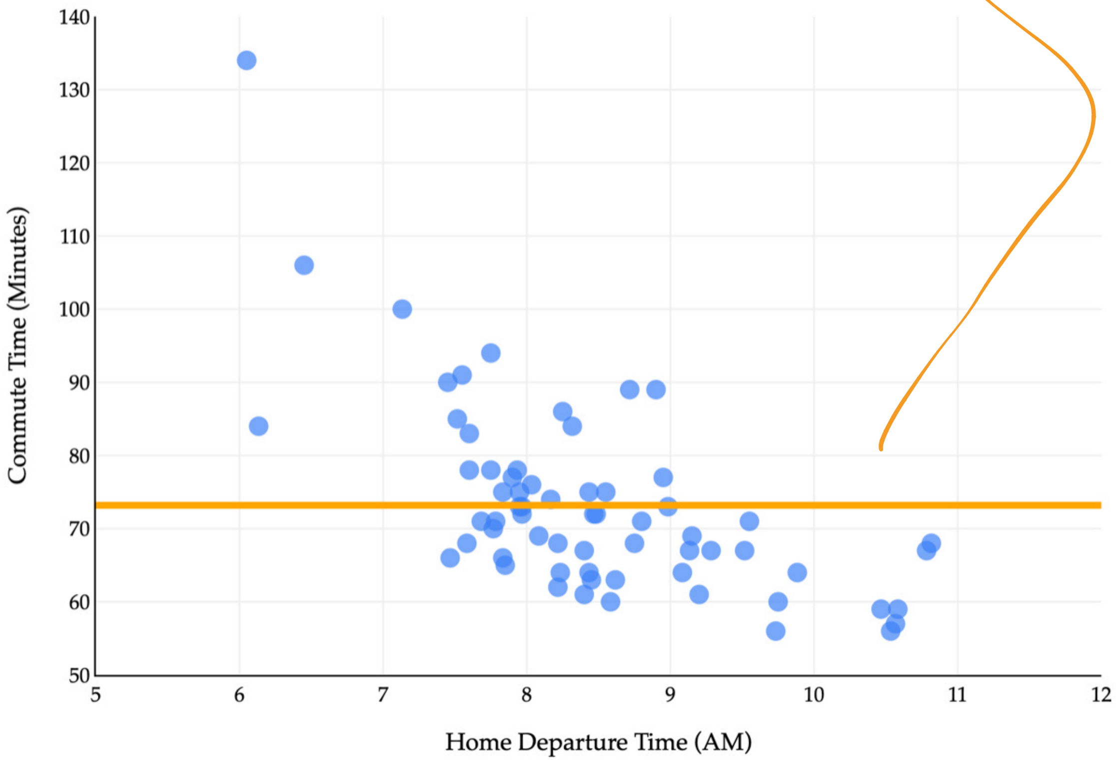
expect more activities in lecture!

1. Recap: The modeling recipe
2. "Empirical risk"
3. Absolute loss vs. squared loss
   - → Minimizing mean absolute error
   - → Outliers
   - → What about degree-100 loss? Degree-0?
4. Center and spread
5. Towards simple linear regression } 1.4 coming soon

Ch. 1.3

read!

Last class:
how do we find the "best" position for this constant model?



Three-step modeling recipe

① Choose a model

constant model

$$h(x_i) = w$$

② Choose a loss function

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

③ Minimize average loss to find optimal parameters

③

$$\frac{1}{n}\sum_{i=1}^{n}\Big(y_i - \underbrace{h(x_i)}\Big)^2$$

constant $\quad h(x_i) = w$ ← parameter

$\downarrow$

$$R_{sq}(w) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i - w)^2}_{\text{function of } w \text{ only!}}$$

optimal parameter

$\downarrow$

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

"$\underbrace{\text{Empirical}}_{\text{comes from data}}$ $\underbrace{\text{risk}}_{\text{fancy term for average loss}}$"

three-step modeling recipe

"empirical risk minimization"

for squared loss, the following are equivalent:

1. "average squared loss"
2. "mean squared error"
3. "empirical risk"

Activity

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

For the constant model, $h(x_i) = w$,
what value of w^*
minimizes average Egg loss?

$$L_E(y_i, w) = (4y_i - 3w)^2$$

$$R_E(w) = \frac{1}{n} \sum_{i=1}^{n} (4y_i - 3w)^2$$ average Egg loss

$$\downarrow$$

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

$$\underbrace{= -\frac{6}{n} \sum_{i=1}^{n} (4y_i - 3w) = 0}_{\text{solve for } w^*}$$

(continued)

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

$$4\sum_{i=1}^{n} y_i - \underbrace{\sum_{i=1}^{n} 3w}_{3w + 3w + \cdots + 3w} = 0$$

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

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

optimal constant

$$\Rightarrow \quad w^* = \frac{4\sum_{i=1}^{n} y_i}{3n} = \boxed{\frac{4}{3}\bar{y}}$$

Absolute loss

Recipe

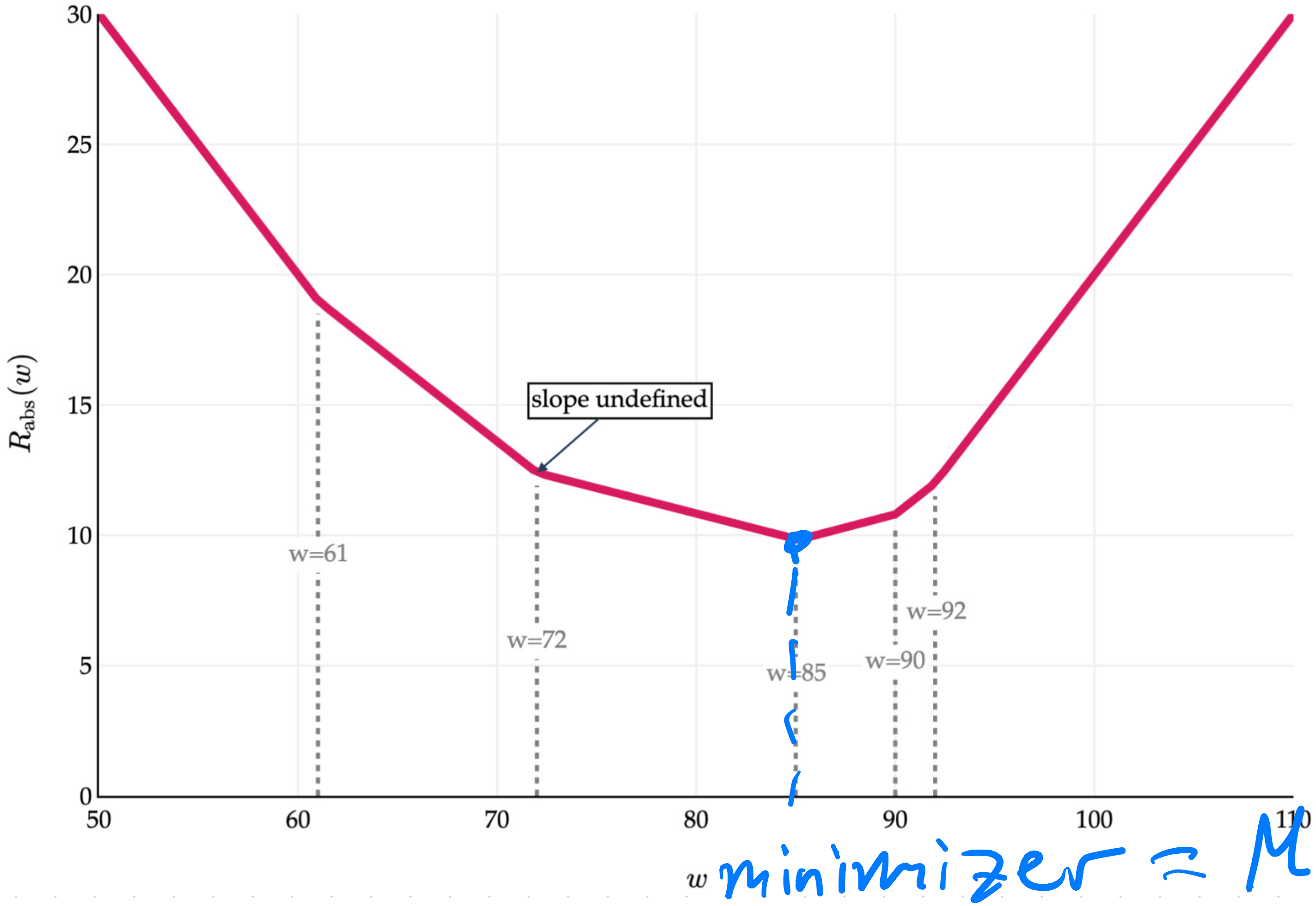
1. $h(x_i) = w$
2. $L_{abs}(y_i, h(x_i)) = |y_i - h(x_i)|$
3. $R_{abs}(w) = \frac{1}{n} \sum_{i=1}^{n} |y_i - w|$

piecewise linear function

"average abs loss" "mean absolute error"

"empirical risk"

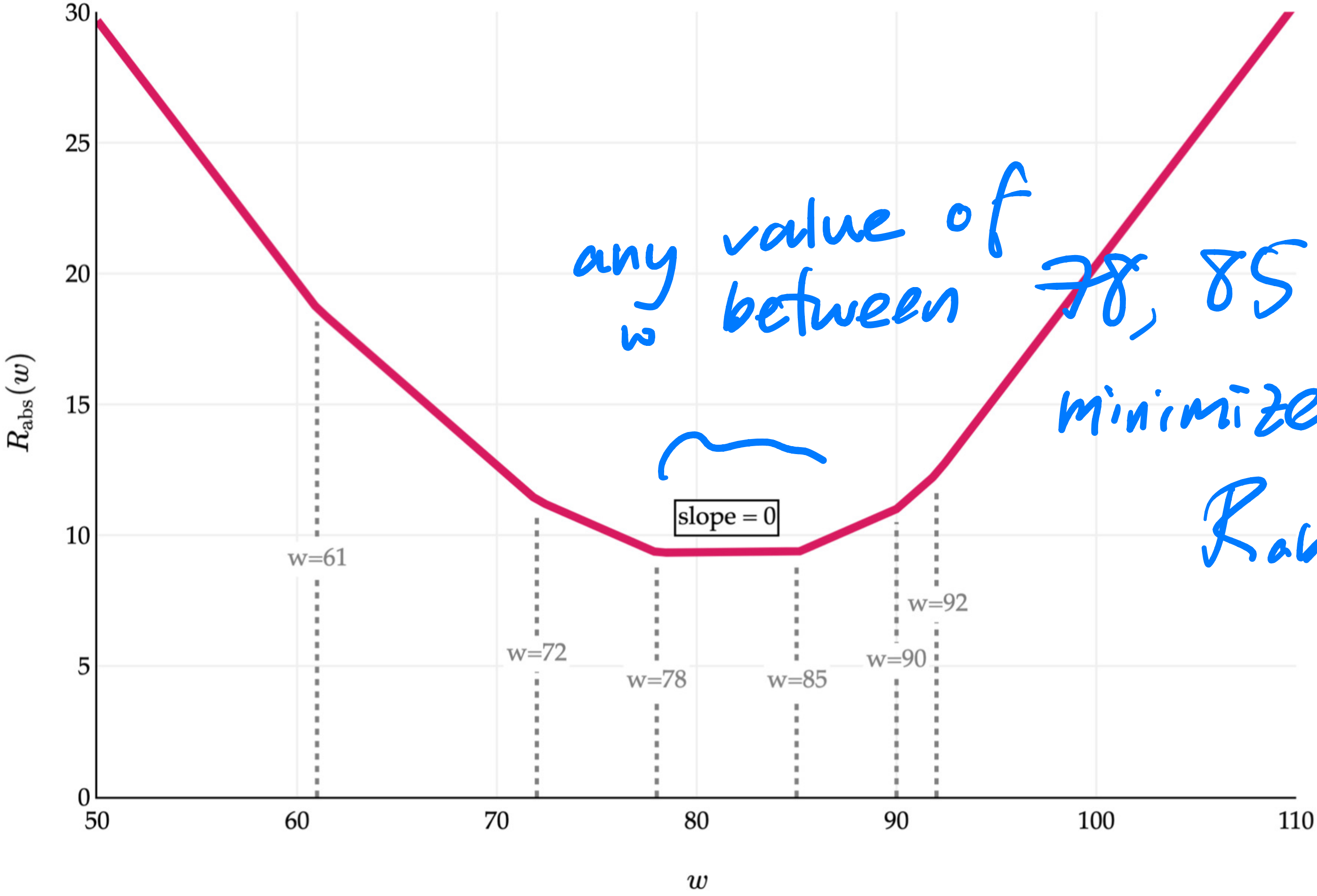
$$R_{\text{abs}}(w) = \frac{1}{5}(|72 - w| + |90 - w| + |61 - w| + |85 - w| + |92 - w|)$$



minimizer = Median

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

And its graph is:



any value of w between 78, 85 minimizes $R_{abs}(w)$

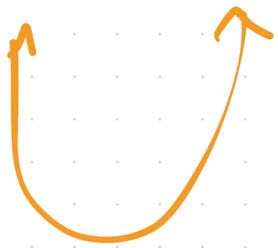
slope = derivative

Goal: Minimize

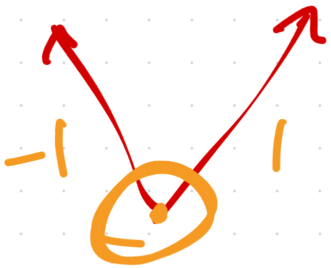
$$R_{abs}(w) = \frac{1}{n}\sum_{i=1}^{n} |y_i - w)$$

try to take $\frac{d}{dw}$

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



$$|y_i - w| = \begin{cases} y_i - w & \text{if } w \le y_i \\ \underbrace{-(y_i - w)}_{\text{outputs}} & \underbrace{\text{if } w > y_i}_{\text{conditions}} \end{cases}$$



$$\frac{d}{dw}|y_i - w| = \begin{cases} -1 & \text{if } w < y_i \\ \text{undefined} & \text{if } w = y_i \\ 1 & \text{if } w > y_i \end{cases}$$

$$\underbrace{\frac{d}{dw} R_{abs}(w)}_{\text{undefined if } w = \text{a data point}} = \frac{1}{n} \sum_{i=1}^{n} \frac{d}{dw} |y_i - w|$$

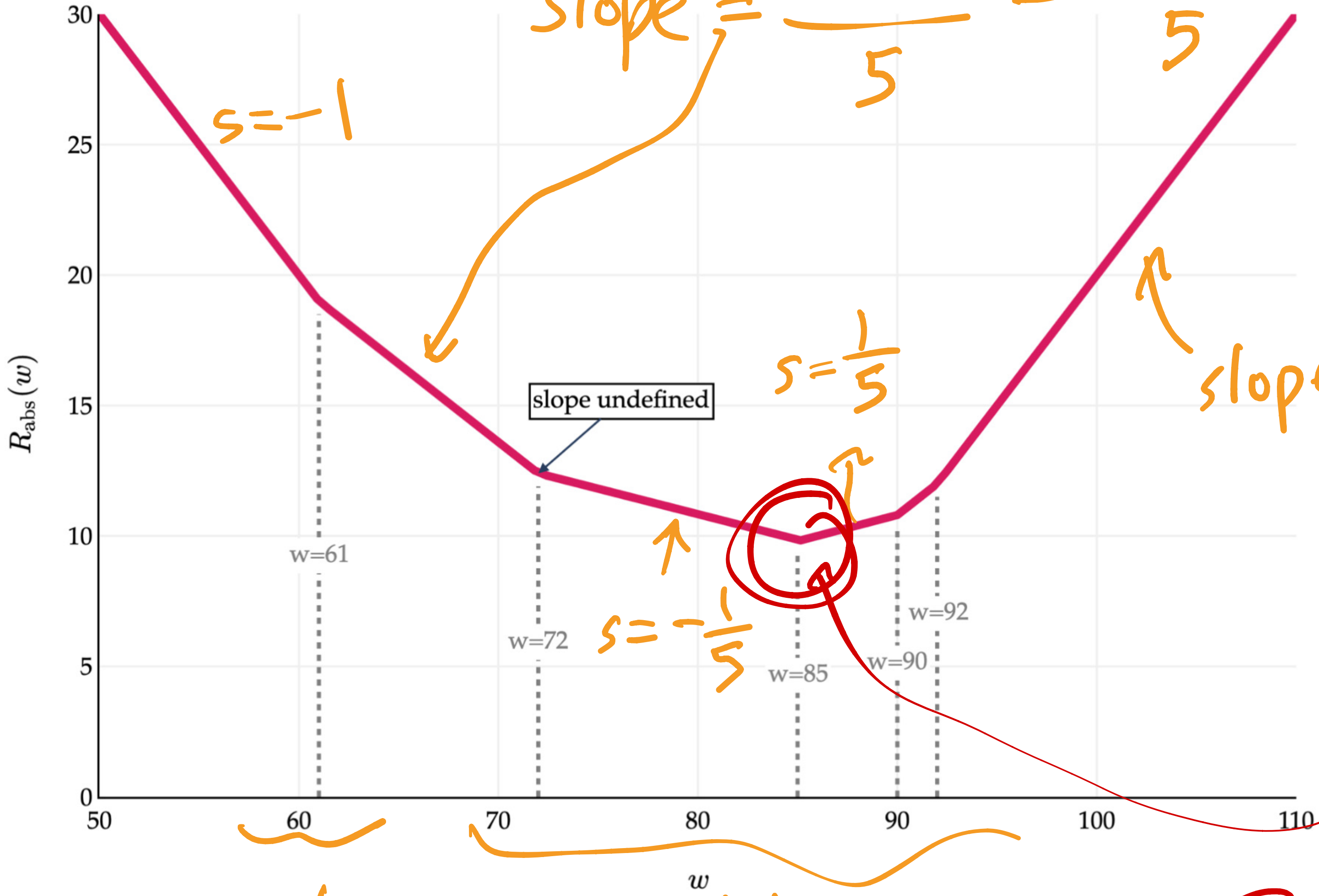
$$= \frac{1}{n} \sum_{i=1}^{n} \underbrace{\begin{cases} -1 & \text{if } w < y_i \\ \text{undefined} & \text{if } w = y_i \\ 1 & \text{if } w > y_i \end{cases}}_{\text{sum of 1s and -1s}}$$

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

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

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

$$\text{slope} = \frac{1 - 4}{5} = -\frac{3}{5}$$



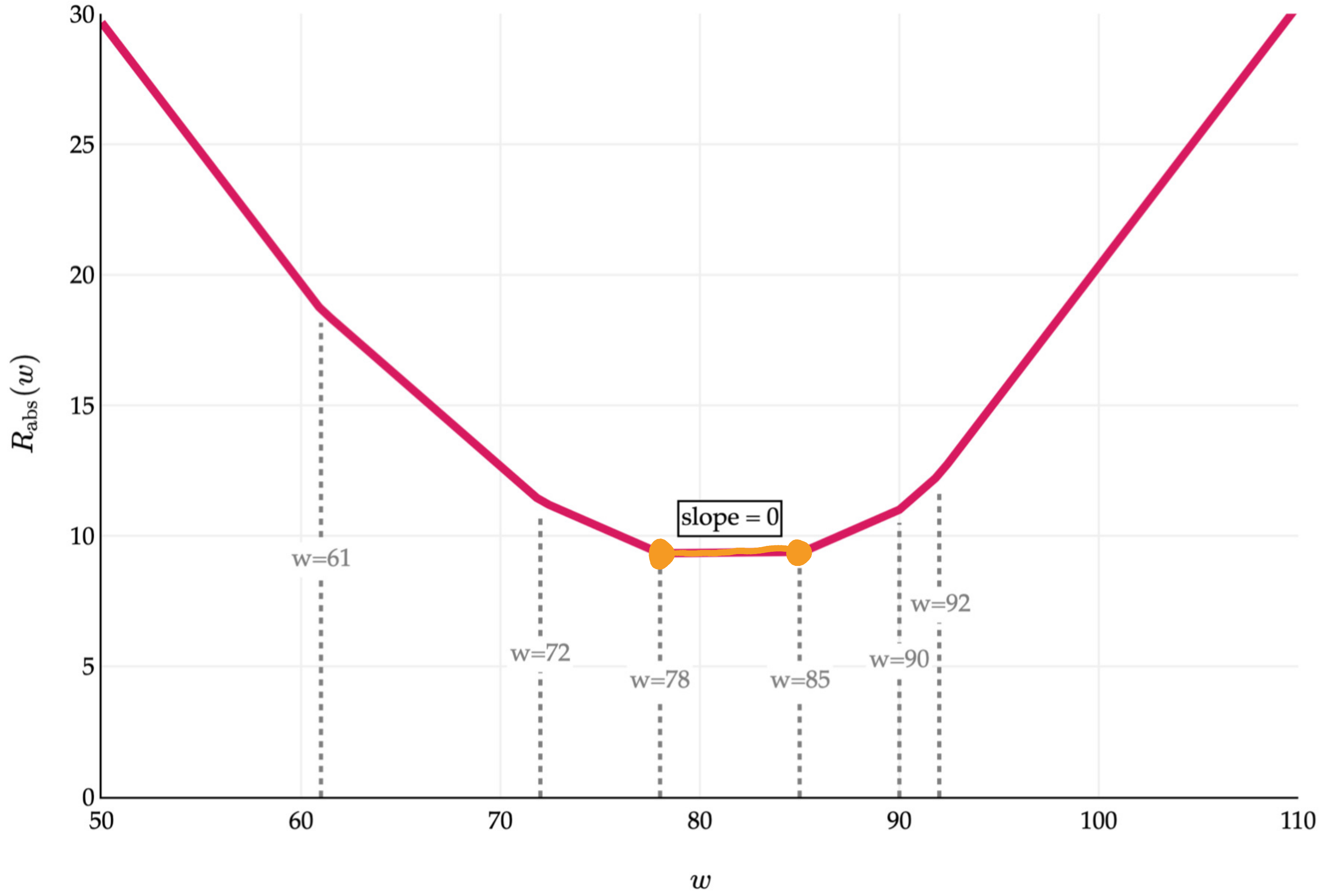
$$\text{slope} = \frac{5 - 0}{5} = 1$$

slope goes from ⊖ → ⊕,

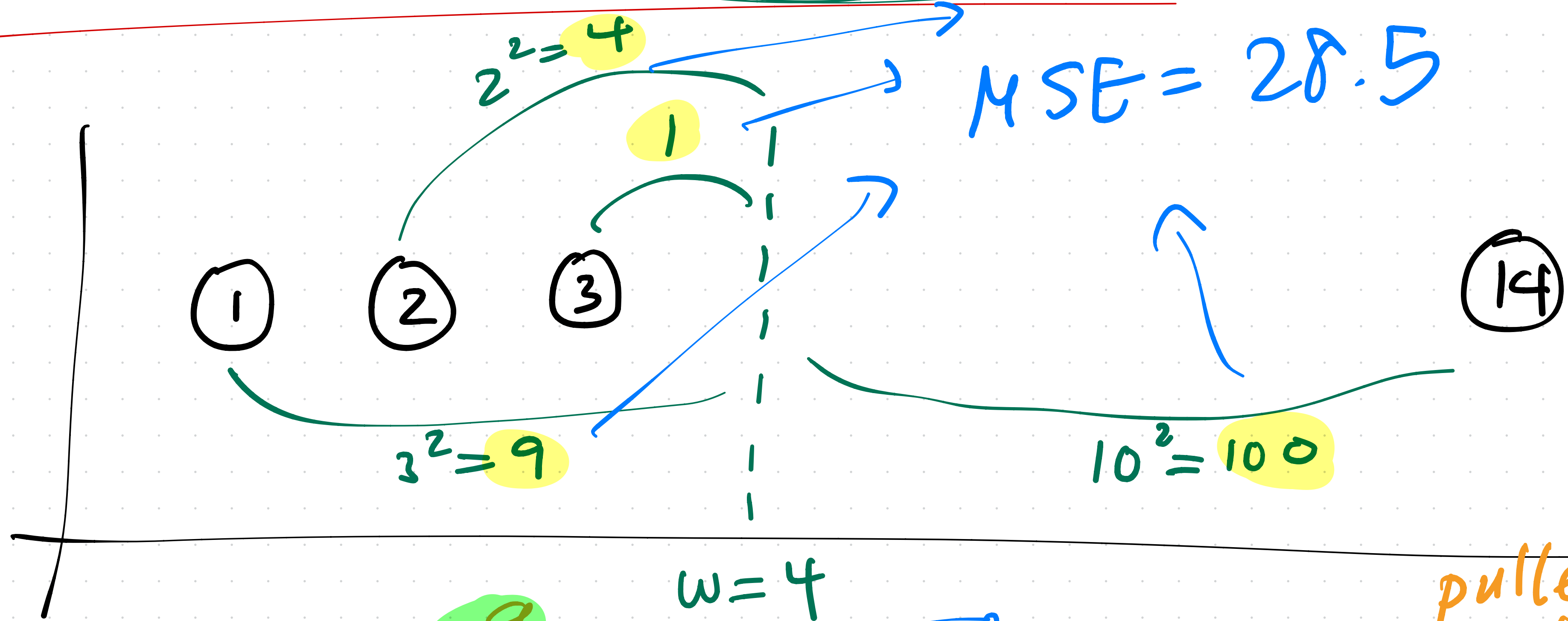
so at bottom

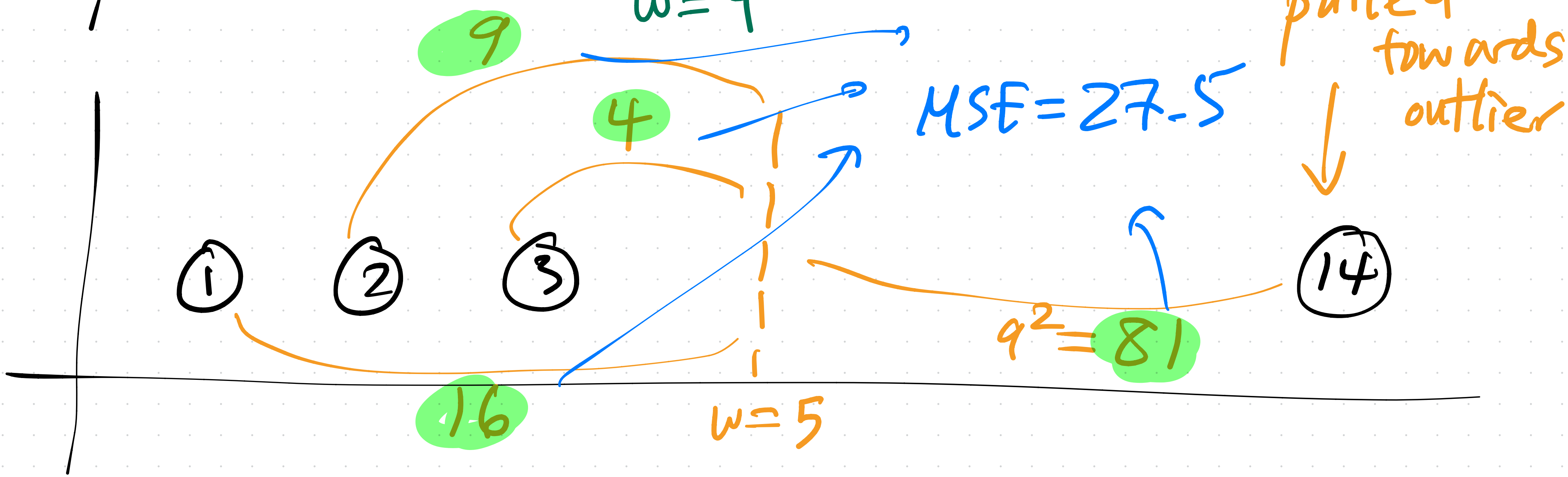
$$R_{\text{abs}}(w) = \frac{1}{6}(|72 - w| + |90 - w| + |61 - w| + |85 - w| + |92 - w| + |78 - w|)$$

And its graph is:



Absolute loss vs. squared loss





"Balance conditions"

① Median

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

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

② Mean

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

$\Rightarrow$ mean : $\sum_{i=1}^{n} (y_i - \text{Mean}) = 0$

e.g. 61 72 85 90 92

mean = 80

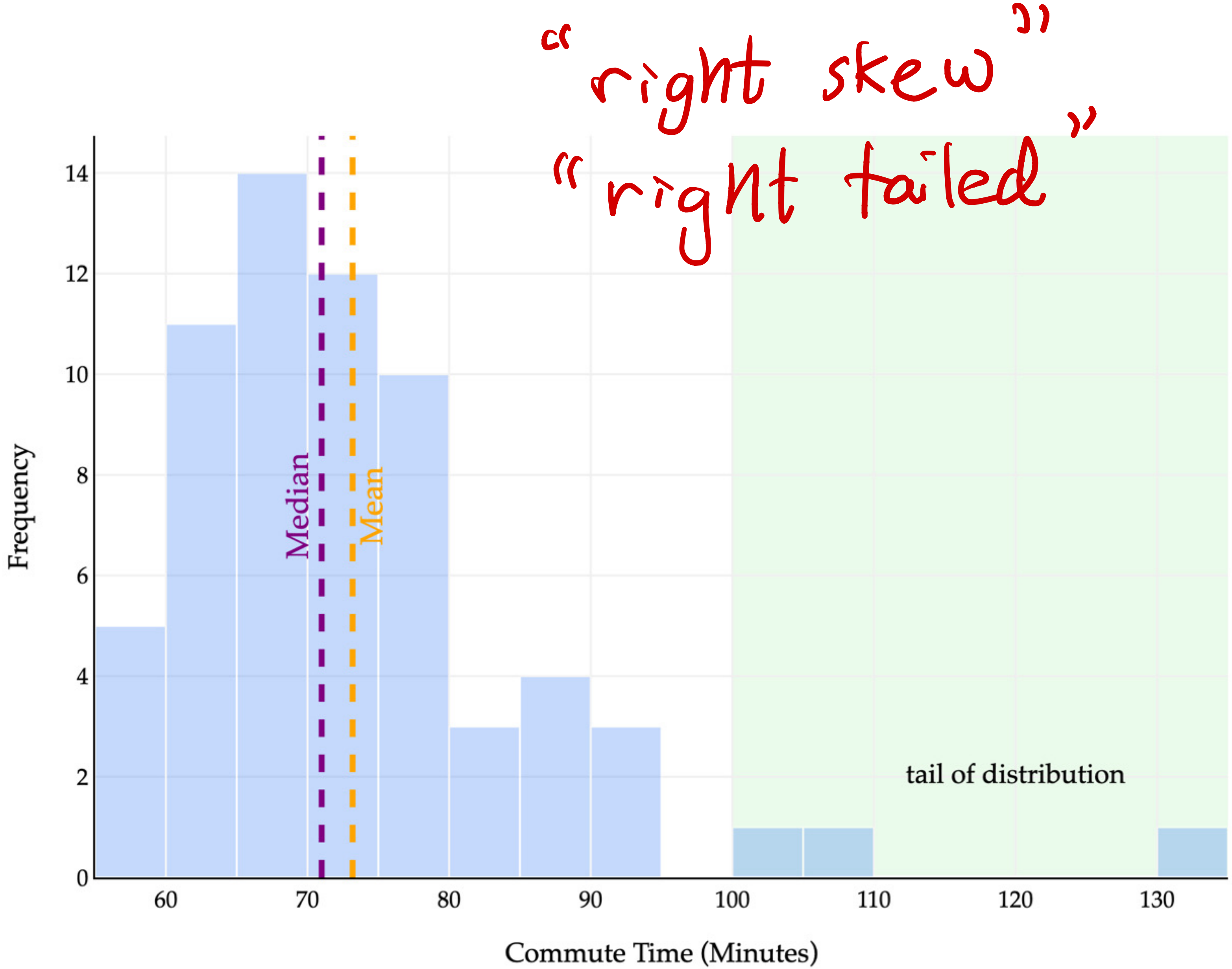
$$\underbrace{61-80}_{-19} \quad \underbrace{72-80}_{-8} \quad \underbrace{85-80}_{5} \quad \underbrace{90-80}_{10} \quad \underbrace{92-80}_{12}$$

$$-19-8=-27 \qquad 5+10+12=27$$

positive deviations

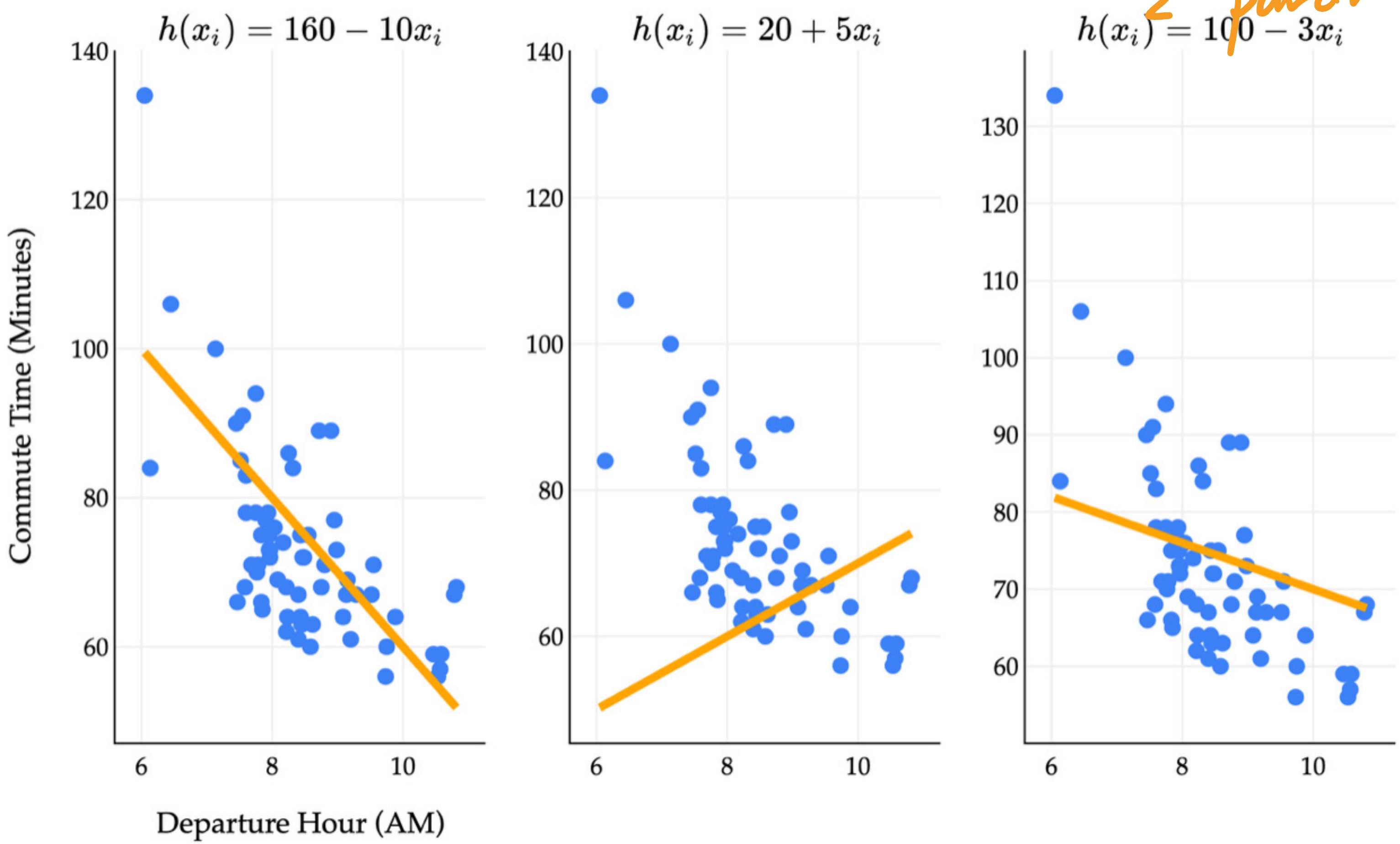
//

negative deviations


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

Simple linear regression

$$h(x_i) = \underbrace{\overset{\text{intercept}}{w_0} + \overset{\text{slope}}{w_1}\, x_i}_{\text{2 parameters}}$$



Recipe

1. $h(x_i) = w_0 + w_1 x_i$ "simple linear"

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

3. minimize average loss

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

Preview:

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

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

SD of y

"correlation coefficient"

$-1 \le r \le 1$

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

depends on first answer