EECS 245, Winter 2026

LEC 2 Models and Loss Functions

→ Read Ch. 1.2-1.3

Agenda

- Terminology: hypothesis functions and parameters
- Loss functions
- Finding optimal parameters using calculus

Ch. 1.2

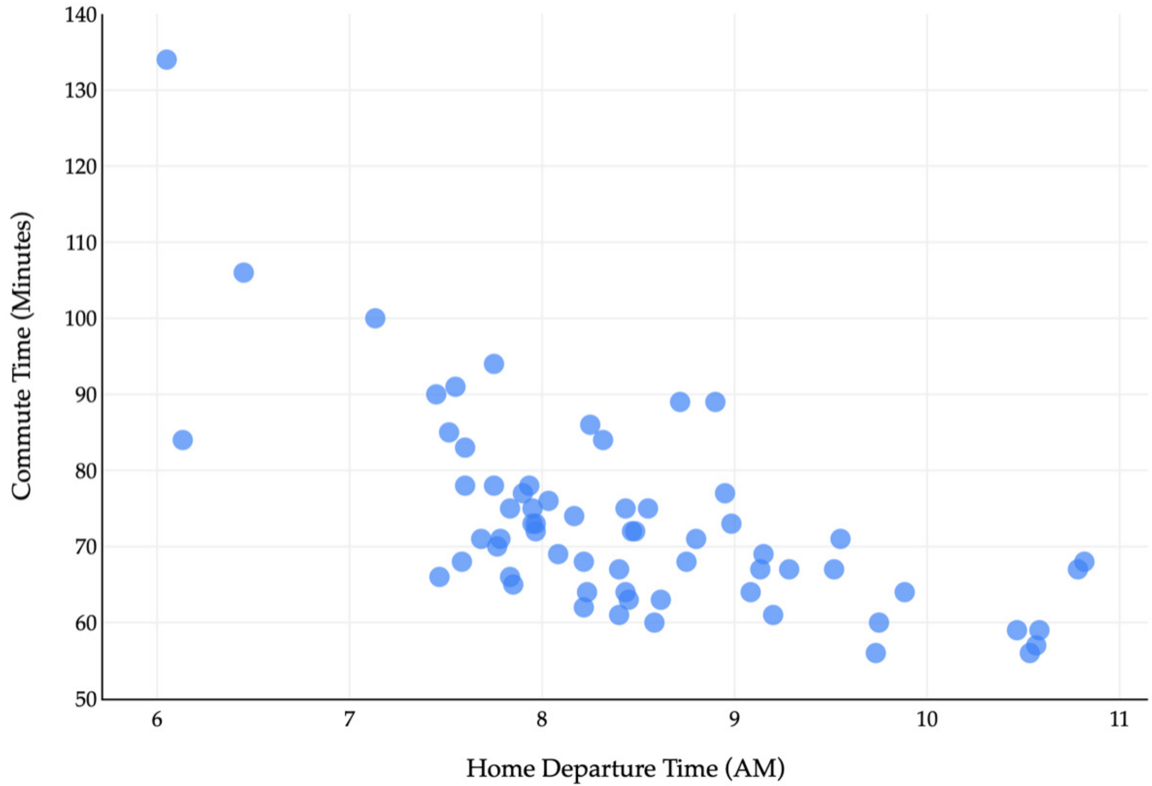
- Exploring different loss functions

1.3

Announcements

- HW 1 due on **Friday**, including Welcome Survey
- See office hours at eecs245.org/calendar

usually have OH after lecture!

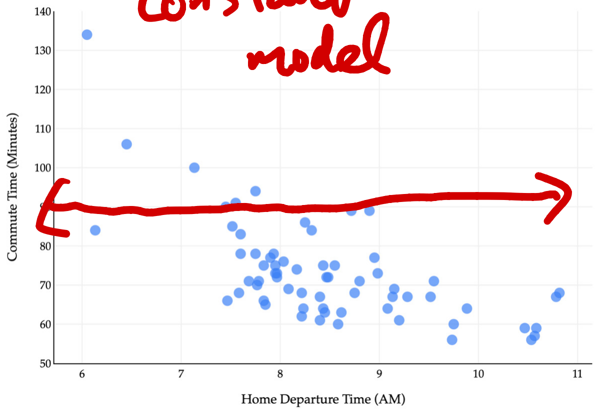

Commute Time (Minutes)
140
130
120
110
100
90
80
70
60
50
6
7
8
9
10
11
Home Departure Time (AM)

Hypothesis function, h, takes in features and outputs predictions y

input/independent x (here: dept time)

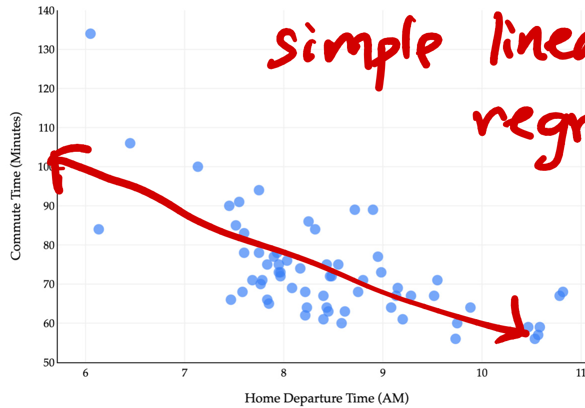
$h(x_i) = 90$

constant model



$h(x_i) = 100 - 4x_i$

simple linear regression



Parameters, w

- Constant model : $h(x_i) = \underbrace{w}$

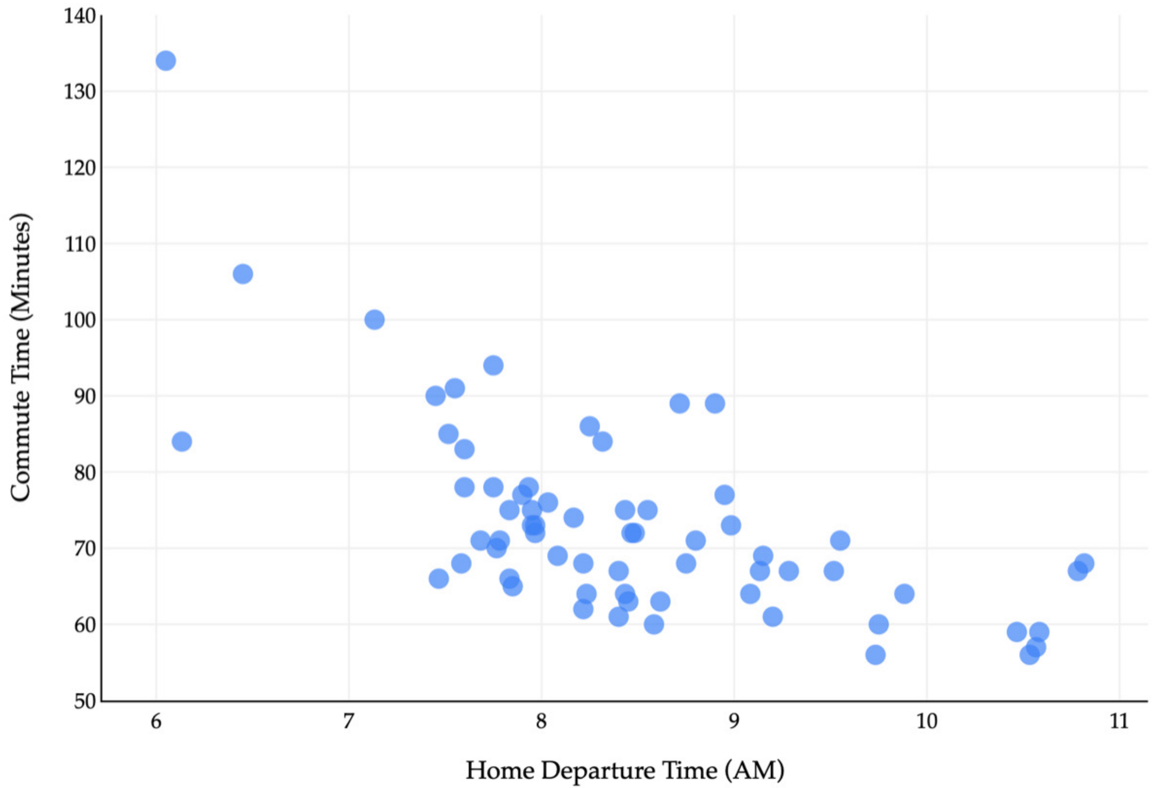
one parameter of the constant model

- Simple linear regression : $h(x_i) = w_0 + w_1 x_i$

intercept

slope

Q: How do we find the best parameters?



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

"Error"

$$\underbrace{e_i}_{} = \underbrace{y_i}_{\text{actual}} - \underbrace{h(x_i)}_{\text{predicted}}$$

for the constant model, this is just w

e.g. $y_i = 80$

1. if $h(x_i) = 75 \longrightarrow e_i = 5$
2. if $h(x_i) = 72 \longrightarrow e_i = 8$
3. if $h(x_i) = 100 \longrightarrow e_i = 80 - 100 = -20$

<u>Loss function</u> : describes the quality of a prediction for a <u>single</u> data point

① Squared loss :

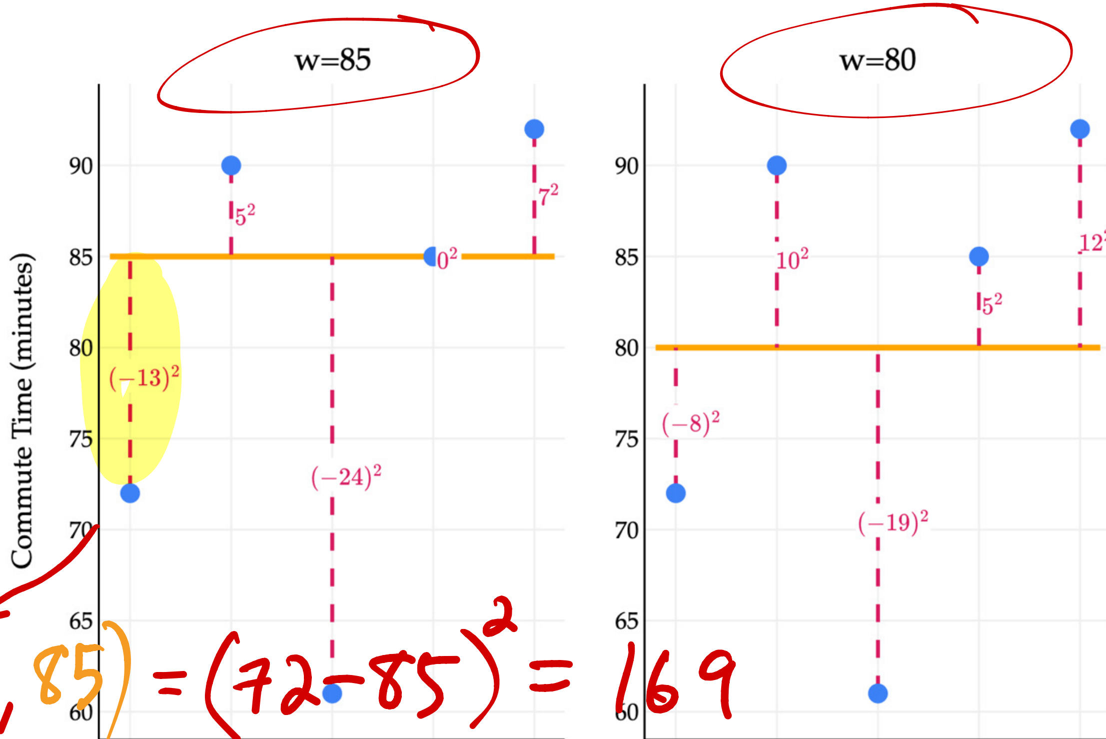
$$L_{sq}(y_i, h(x_i)) = \underbrace{(y_i - h(x_i))^2}_{(\text{actual} - \text{predicted})^2}$$

② Absolute loss :

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

$$\underset{y_1}{72},\ \underset{y_2}{90},\ \underset{y_3}{61},\ \underset{y_4}{85},\ \underset{y_5}{92}$$

← Visualizing squared loss



$$L_{sq}(72, 85) = (72-85)^2 = 169$$

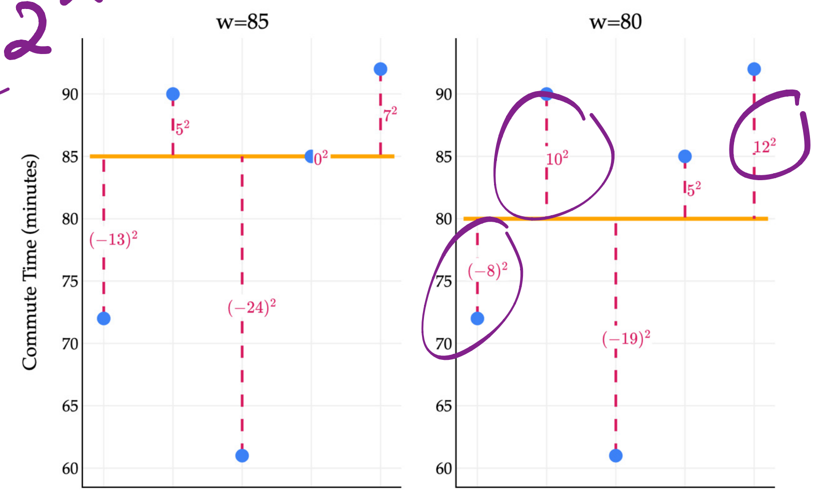
Average squared loss gives me one number that describes the quality of a w for the whole dataset!

- For $w=85$: $\dfrac{(-13)^2+5^2+\dots+7^2}{5} = 163.8$

- For $w=80$: $\dfrac{(-8)^2+10^2+\dots+12^2}{5}$

$= 138.8$

80 is better than 85,
bc $138.8 < 163.8$!

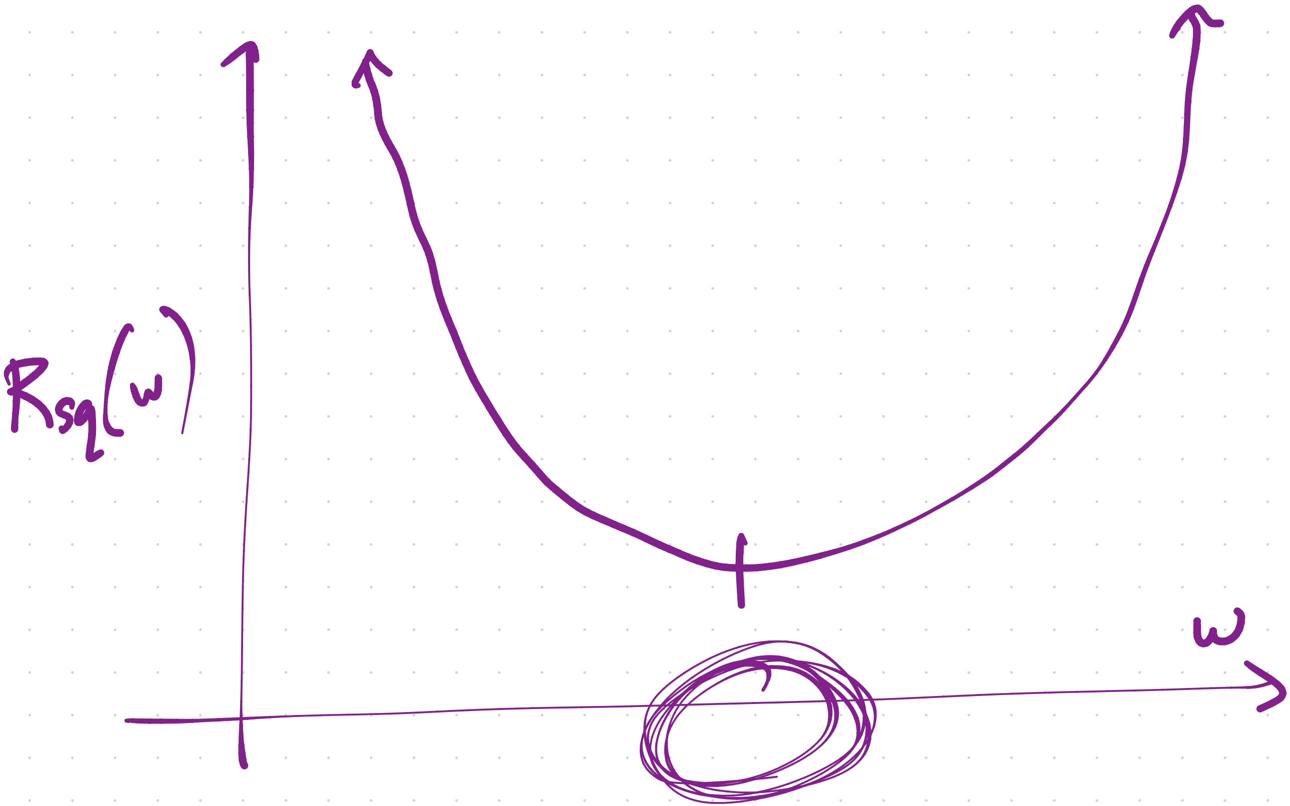


$y_1 = 72, \; y_2 = 90, \; y_3 = 61, \; y_4 = 85, \; y_5 = 92$

$$\underbrace{R_{sq}(w)}_{\text{average squared loss}} = \frac{(72-w)^2 + (90-w)^2 + (61-w)^2 + (85-w)^2 + (92-w)^2}{5}$$

L: loss for a single data point

R: average loss across a whole dataset



$R_{sq}(w)$
w

Given $y_1, y_2, \ldots, y_n$ (all numbers)

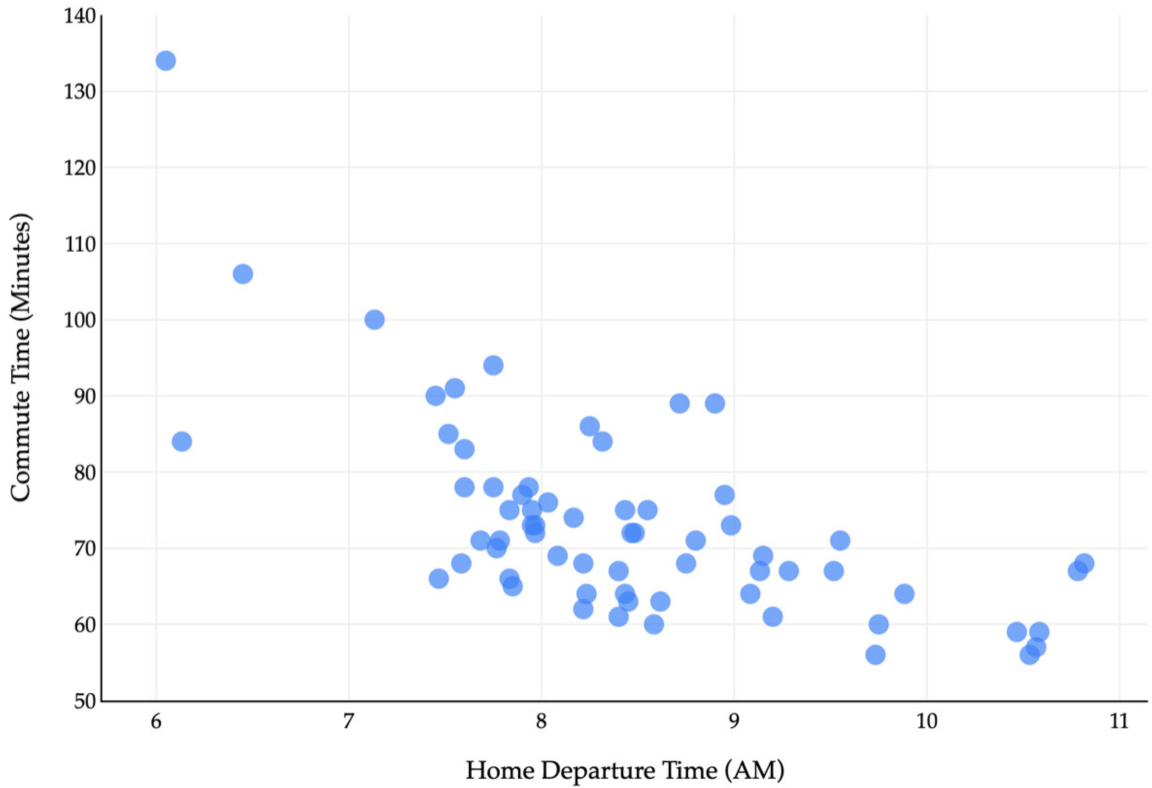
Goal is to find the best constant prediction, w, by minimizing R_{sq} (average squared loss)!

$$R_{sq}(w) = \frac{(y_1 - w)^2 + (y_2 - w)^2 + \cdots + (y_n - w)^2}{n}$$

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

Think of the y_i's as constants; the only "variable" is w!

"average squared loss" "mean squared error"


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

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

1. Take derivative wrt w
2. set to 0 and solve
3. second derivative test

1

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

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

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

$$\frac{dR}{dw}$$

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

set to 0

Aside: what is

$$\frac{d}{dw}(y_i - w)^2$$

$$= 2(y_i - w)\underbrace{\frac{d}{dw}(y_i - w)}_{\text{chain rule}}$$

$$= 2(y_i - w)(-1)$$

$$= -2(y_i - w) \quad = 2(w - y_i)$$

②

$$-\frac{2}{n}\sum_{i=1}^{n}(y_i - w) = 0$$

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

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

best/optimal parameter

mean of y_i's!

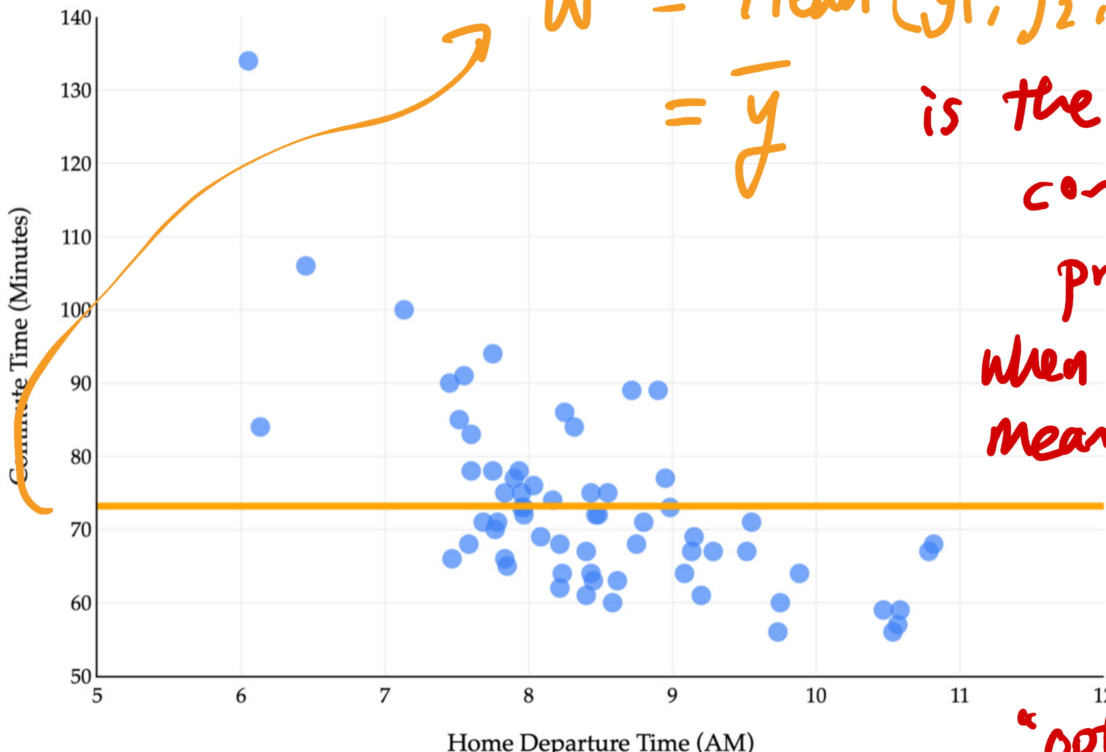
$$\sum_{i=1}^{n} y_i - nw = 0 \Rightarrow \boxed{w^* = \frac{\sum_{i=1}^{n} y_i}{n}}$$

③ second derivative test

$\longrightarrow$ Read Ch. 1.2

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

is the optimal constant prediction, when minimizing mean squared error



w^*

"optimal model parameter"

1.2

1.3

"Three-step modeling recipe" for making predictions

① Choose a model

$$h(x_i) = w$$ "constant model"

② Choose a loss function

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

③ Minimize <u>average</u> loss to find optimal parameters

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

What if we use absolute loss and the constant model?

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