EECS 245, Spring 2026

LEC 1 — Introduction, Models, and Loss Functions

→ Read: Ch. 1 of notes

→ Course website: eecs245.org

Agenda

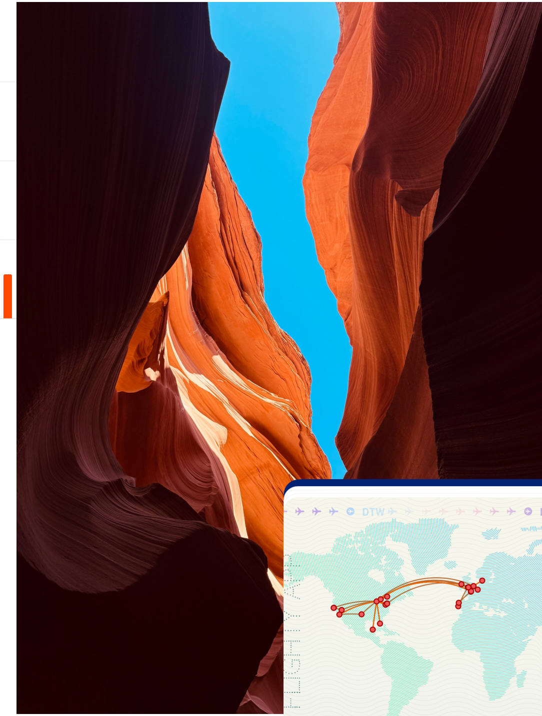
- Introductions
- What is machine learning, and what will we learn in this class? ← Ch. 1.1
- Logistics
- Models and loss functions ← Ch. 1.2 - 1.3
- Finding optimal parameters

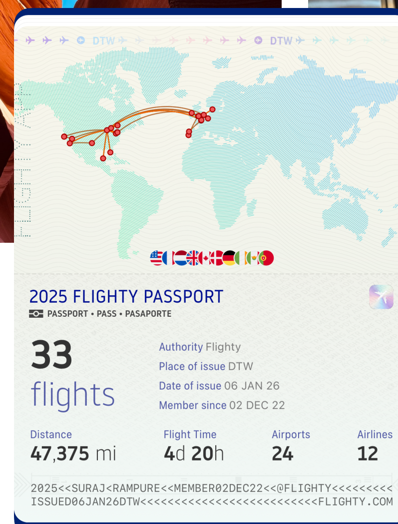
Who am I?

* Call me Suraj "soo-rudge"

- Just finished 2nd year as faculty @ Michigan
- Taught data science at UC San Diego 2021-24
- BS and MS in EECS from UC Berkeley
- From Windsor, ON 🇨🇦
- No TAs this spring – just me


TOTAL
36,975 steps
Yesterday
6



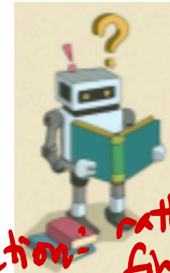

DTW
2025 FLIGHTY PASSPORT
PASSPORT • PASS • PASAPORTE
33
flights
Authority Flighty
Place of issue DTW
Date of issue 06 JAN 26
Member since 02 DEC 22
Distance
47,375 mi
Flight Time
4d 20h
Airports
24
Airlines
12
2025<<SURAJ<RAMPURE<<MEMBER02DEC22<<@FLIGHTY<<<<<<<<<
ISSUED06JAN26DTW<<<<<<<<<<<<<<<<<<<<<<<<<<FLIGHTY.COM

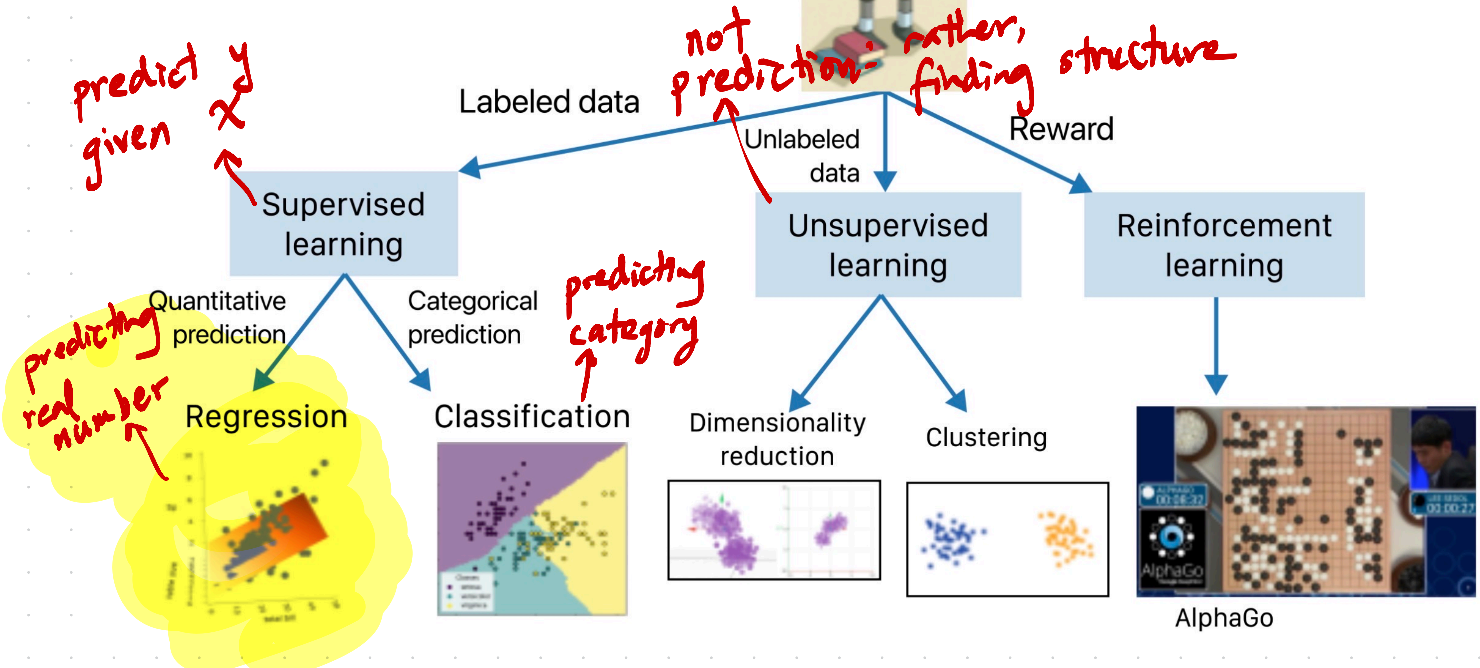

HAPPY BIRTHDAY

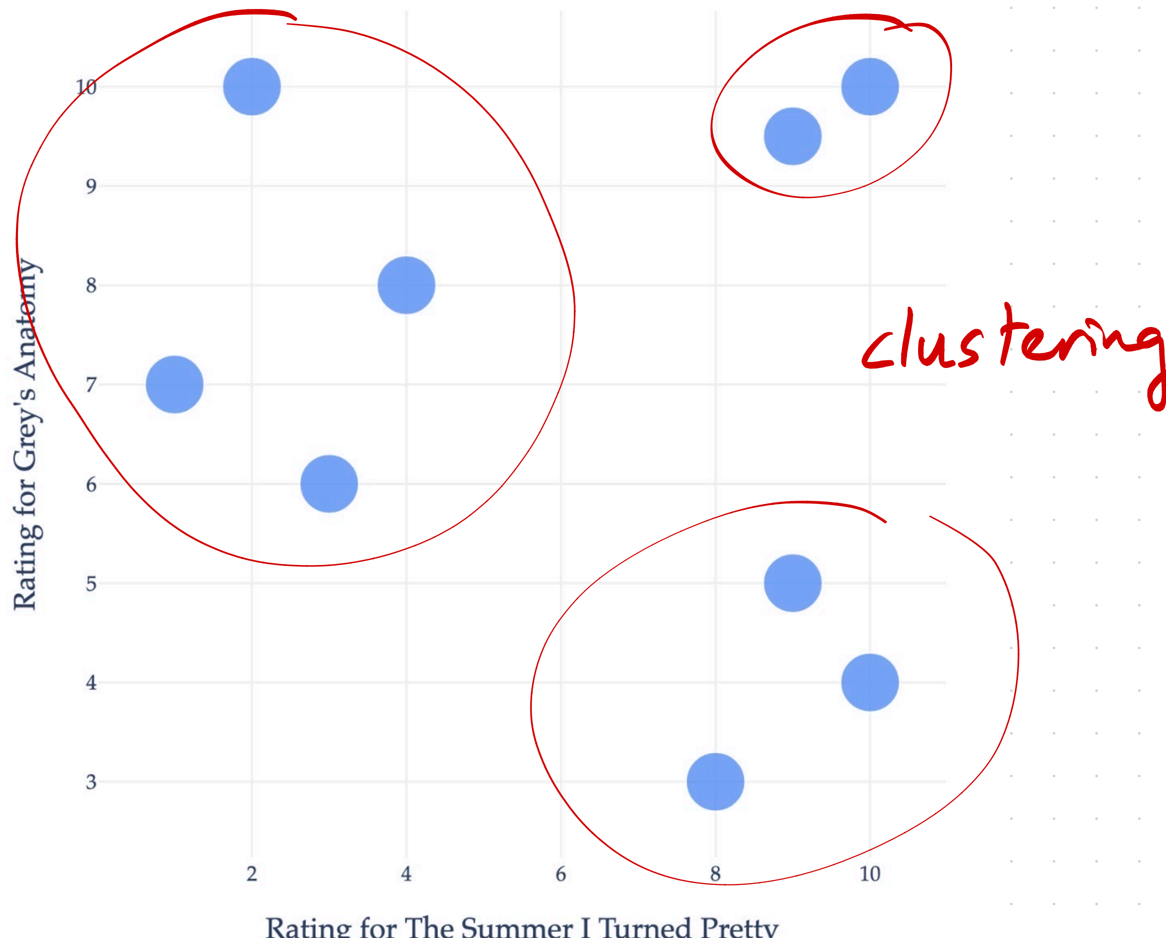
Machine learning: automatically

learning patterns from data

Taxonomy of machine learning



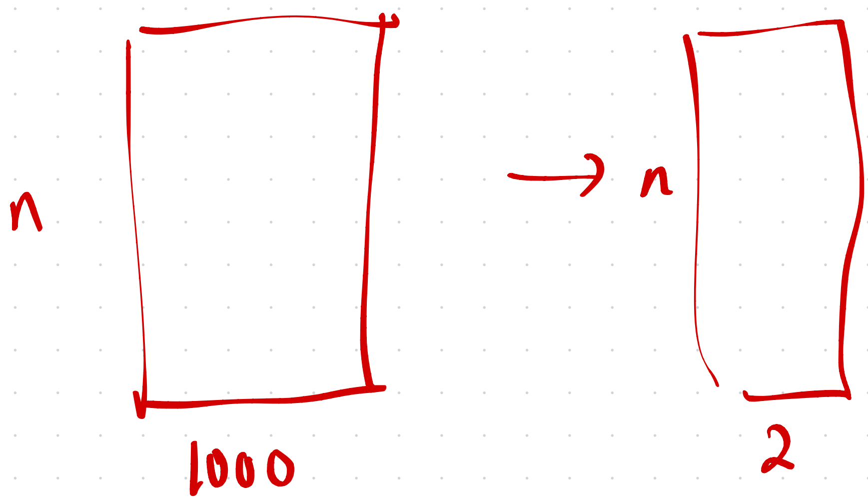




Rating for Grey's Anatomy
10
9
8
7
6
5
4
3
2
4
6
8
10
Rating for The Summer I Turned Pretty
clustering

another example of unsupervised learning:

dimensionality reduction



Math for ML

1. Introduction to Supervised Learning
2. Simple Linear Regression
3. Vectors
4. Linear Independence

Midterm 1

5. Matrices
6. Linear Transformations and Projections
7. Regression using Linear Algebra
8. Gradients

Midterm 2

9. Eigenvalues and Eigenvectors
10. Singular Value Decomposition

post-Midterm 2

eecs245.org

starting Ch. 1.2 →

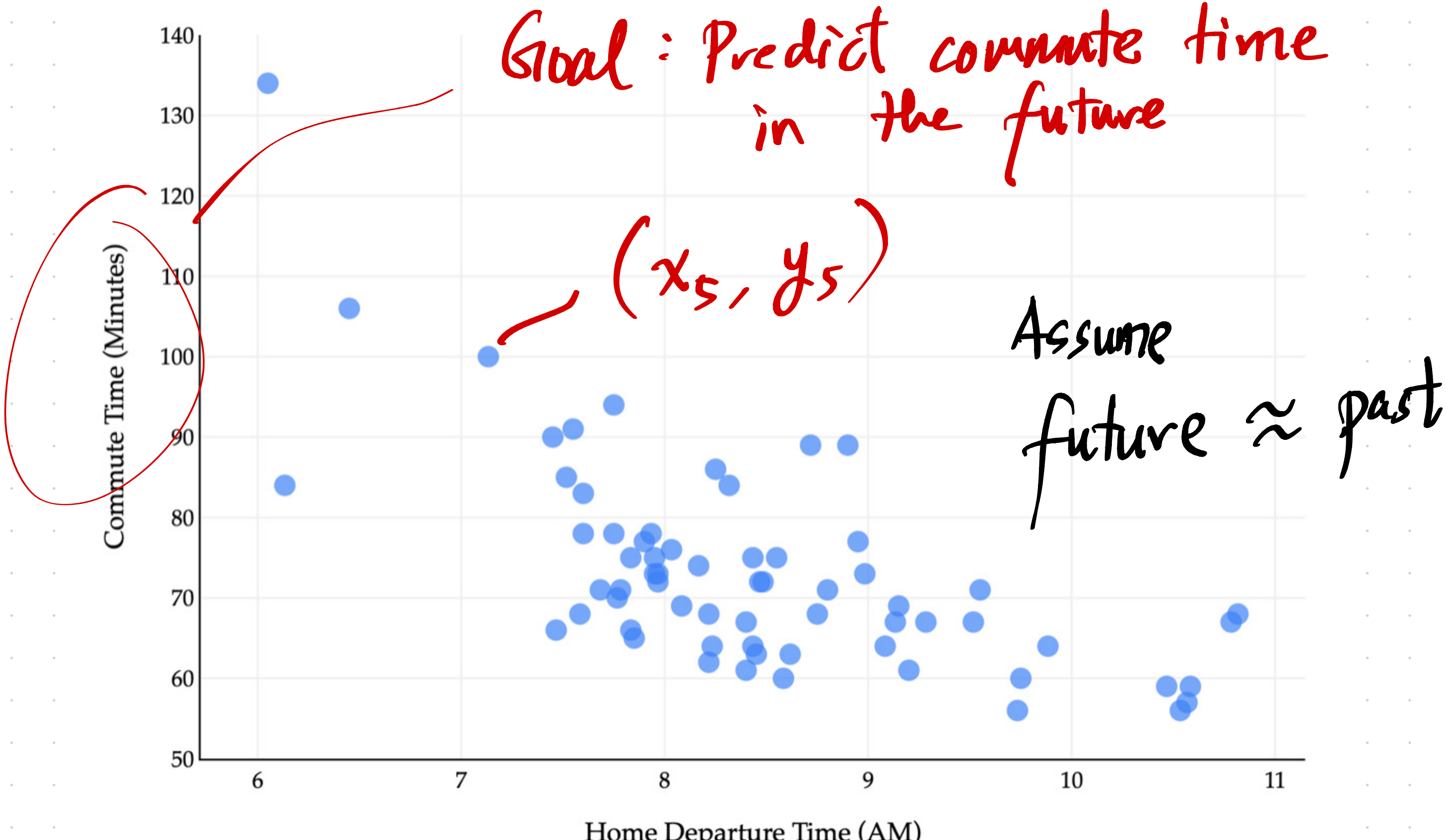
	date	day	departure_hour	minutes
0	5/15/2023	Mon	10.816667	68.0
1	5/16/2023	Tue	7.750000	94.0
2	5/22/2023	Mon	8.450000	63.0
3	5/23/2023	Tue	7.133333	100.0
4	5/30/2023	Tue	9.150000	69.0

features (date, day); feature (departure_hour); label/target/output (minutes) — commute time

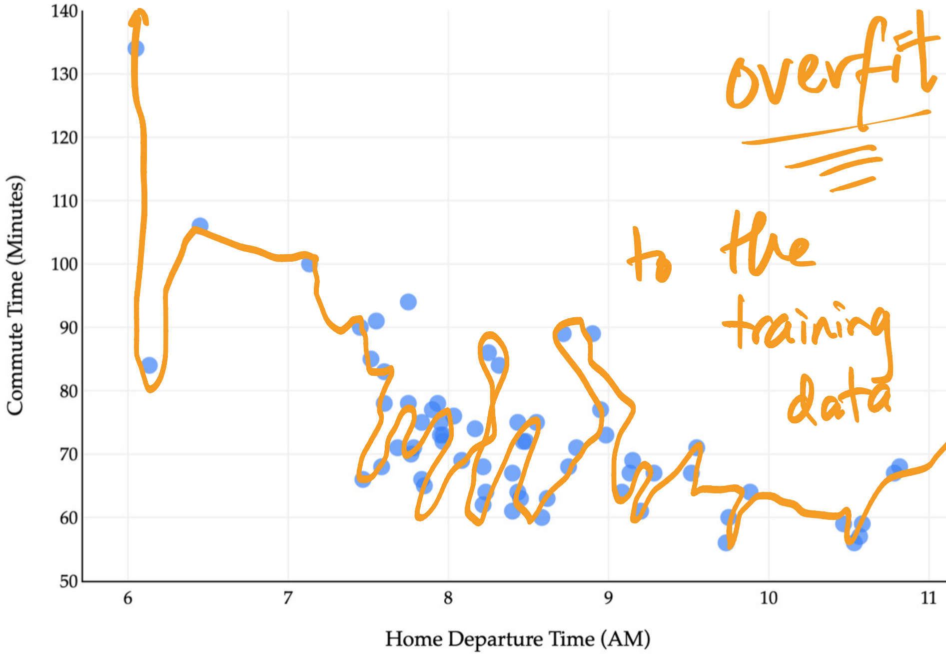
x_2 = 7.750000, y_2 = 94.0 — "y"

8 am + 45% of an hour = 8:27am

Goal: given an x_i predict y_i

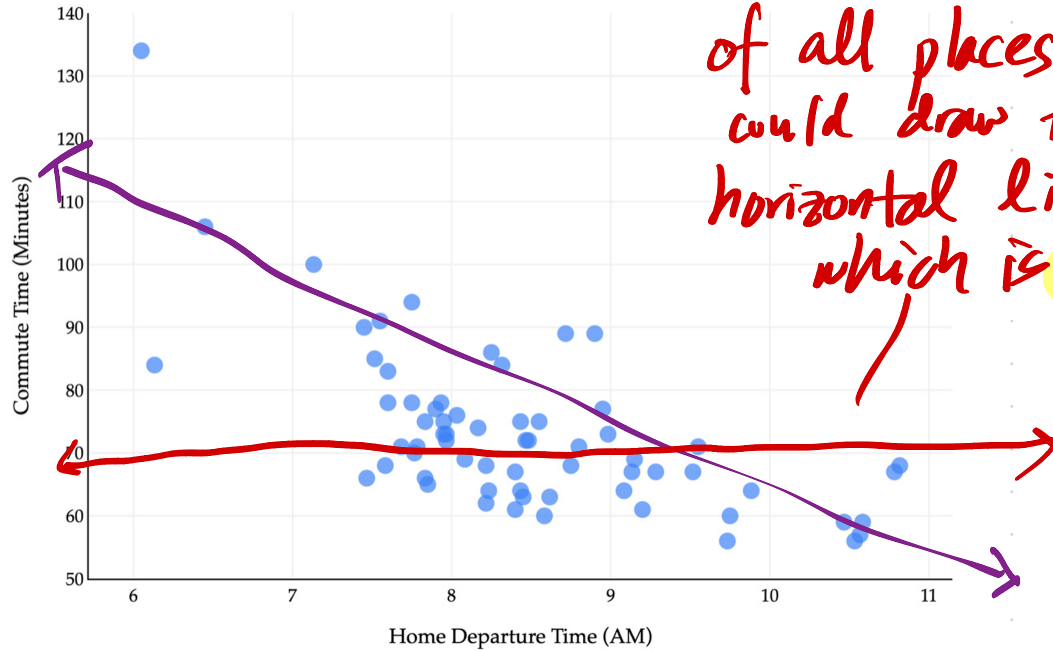

Goal: Predict commute time in the future
(x5, y5)
Assume future ≈ past
Commute Time (Minutes)
140
130
120
110
100
90
80
70
60
50
6
7
8
9
10
11
Home Departure Time (AM)

Model: A set of assumptions about how data was generated



overfit to the training data

bad: we want to generalize to unseen data.



of all places I could draw this horizontal line, which is best?

① constant model

② simple linear regression

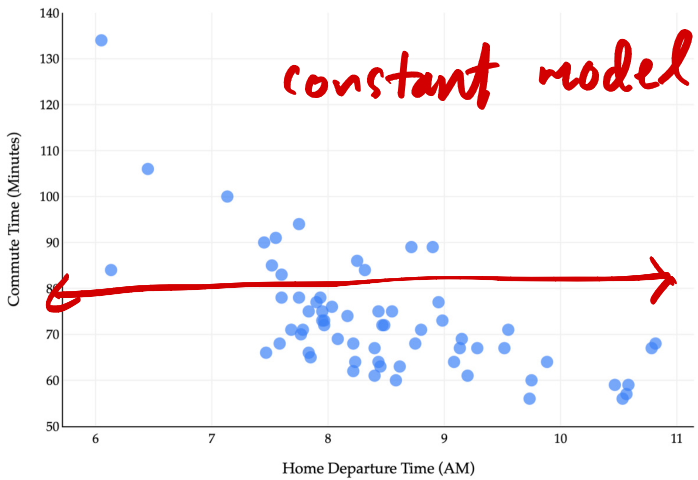
line of best fit

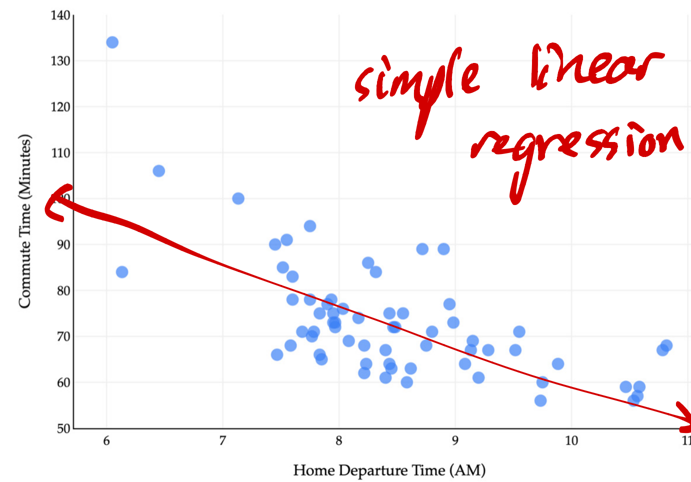
Hypothesis function, h, takes in features and returns predictions (predicted y-values)

departure hour$_i$

$$h(x_i) = 80$$

$$h(x_i) = 100 - 4x_i$$





Parameters, w

Q: How do you find the "best" parameters?

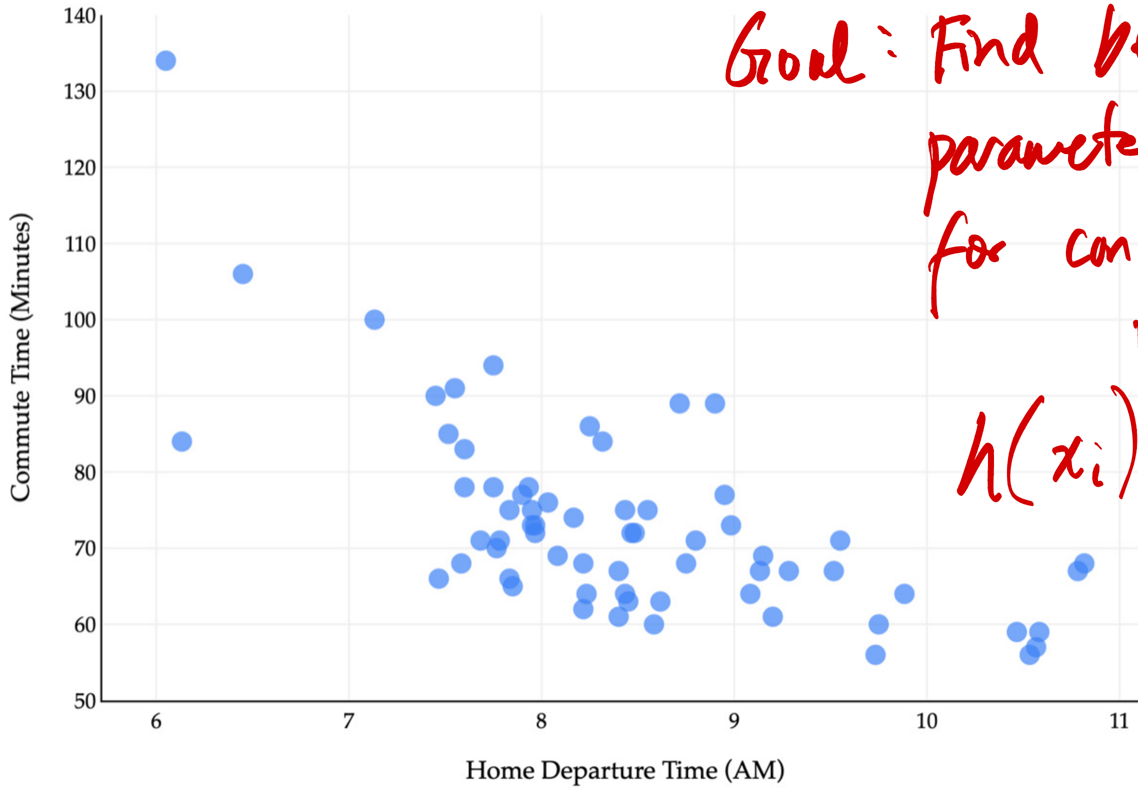
- Constant model: one parameter

$$h(x_i) = w$$

- Simple linear regression model: 2 parameters

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

w_0 ↑ slope w_1 ↑ intercept



Goal: Find best parameter for constant model

$h(x_i) = w$

Error

$$e_i = y_i - h(x_i)$$

y_i: actual

$h(x_i)$: predicted

for constant model, this is just $h(x_i) = w$

e.g. $y_i = 80$

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

Loss function: describes the quality of a prediction for a single data point

① Squared loss:

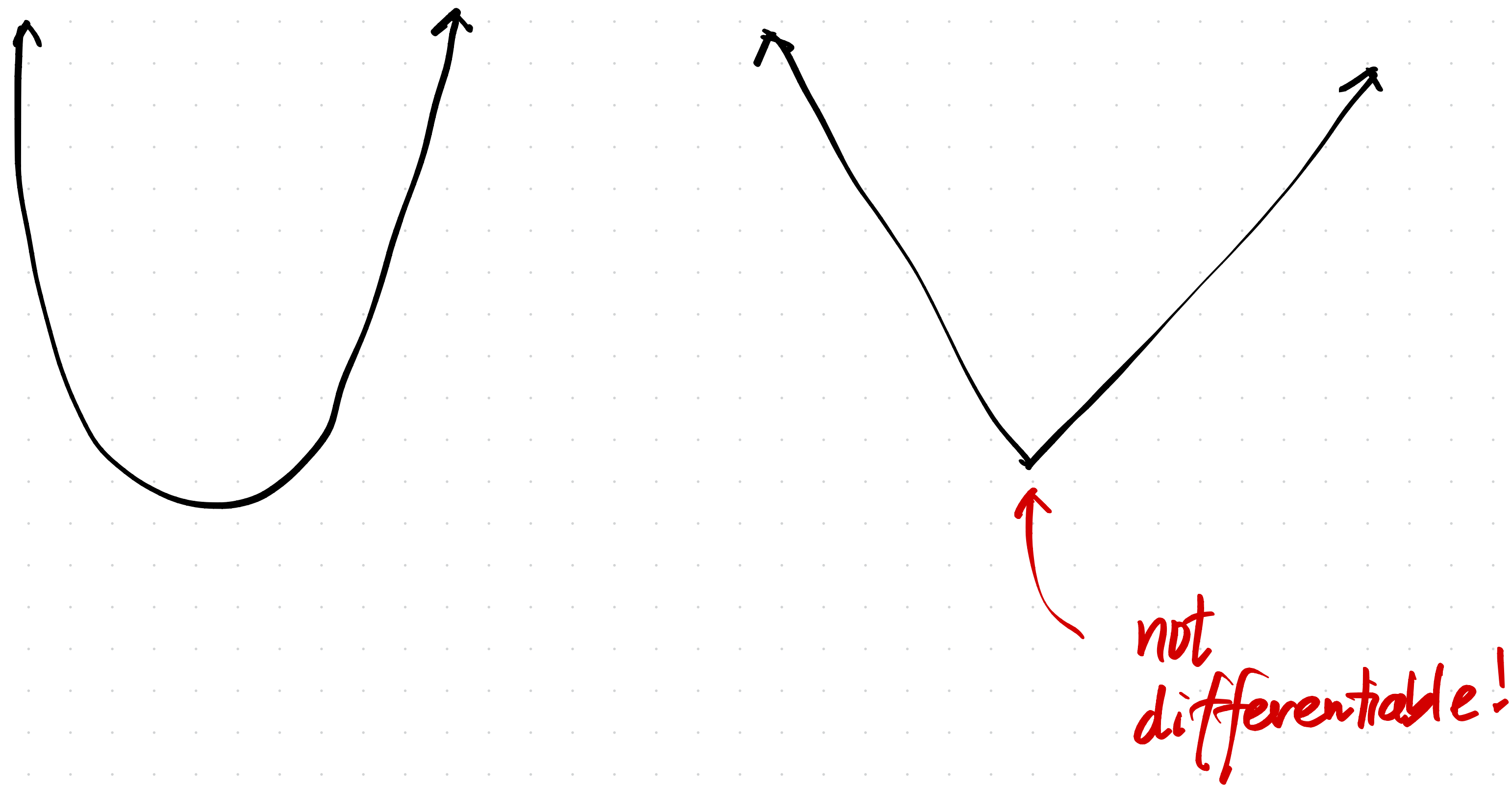
$$L_{sq}(y_i, h(x_i)) = (y_i - h(x_i))^2$$

default choice

(actual − predicted)^2

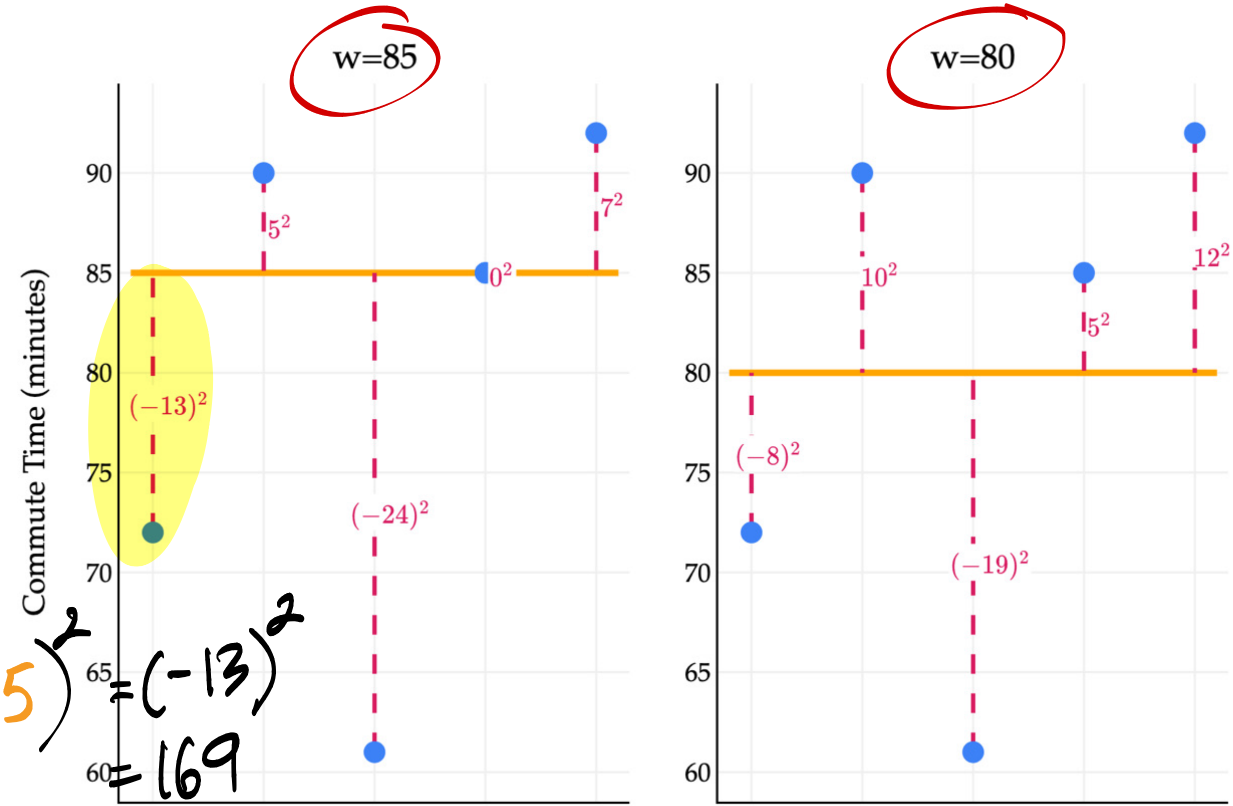
② Absolute loss:

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


not differentiable!

72, 90, 61, 85, 92

y_1 y_2 y_3 y_4 y_5



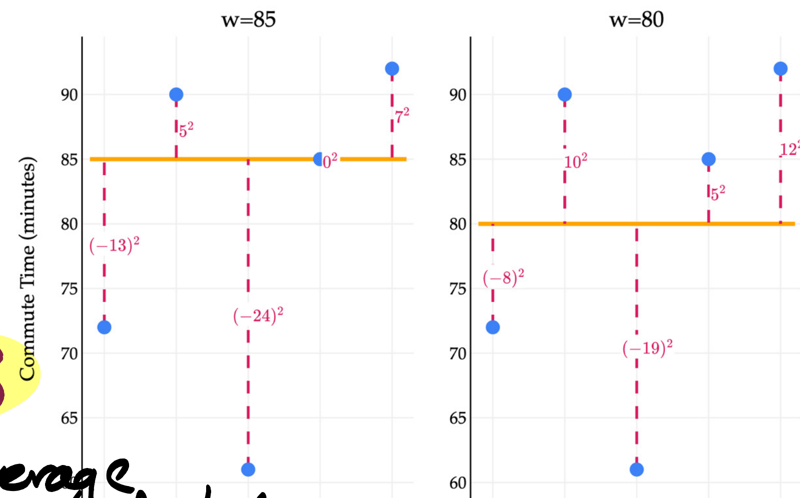
$(72-85)^2 = (-13)^2$

$= 169$

Average squared loss gives **one number** that describes the quality of a parameter across the whole dataset

- $w = 85$: $\dfrac{(-13)^2 + 5^2 + (-24)^2 + 0^2 + 7^2}{5} = 163.8$

- $w = 80 = \dfrac{(-8)^2 + 10^2 + (-19)^2 + 5^2 + 12^2}{5} = 138.8$



80 better than 85: lower average squared loss

72, 90, 61, 85, 92

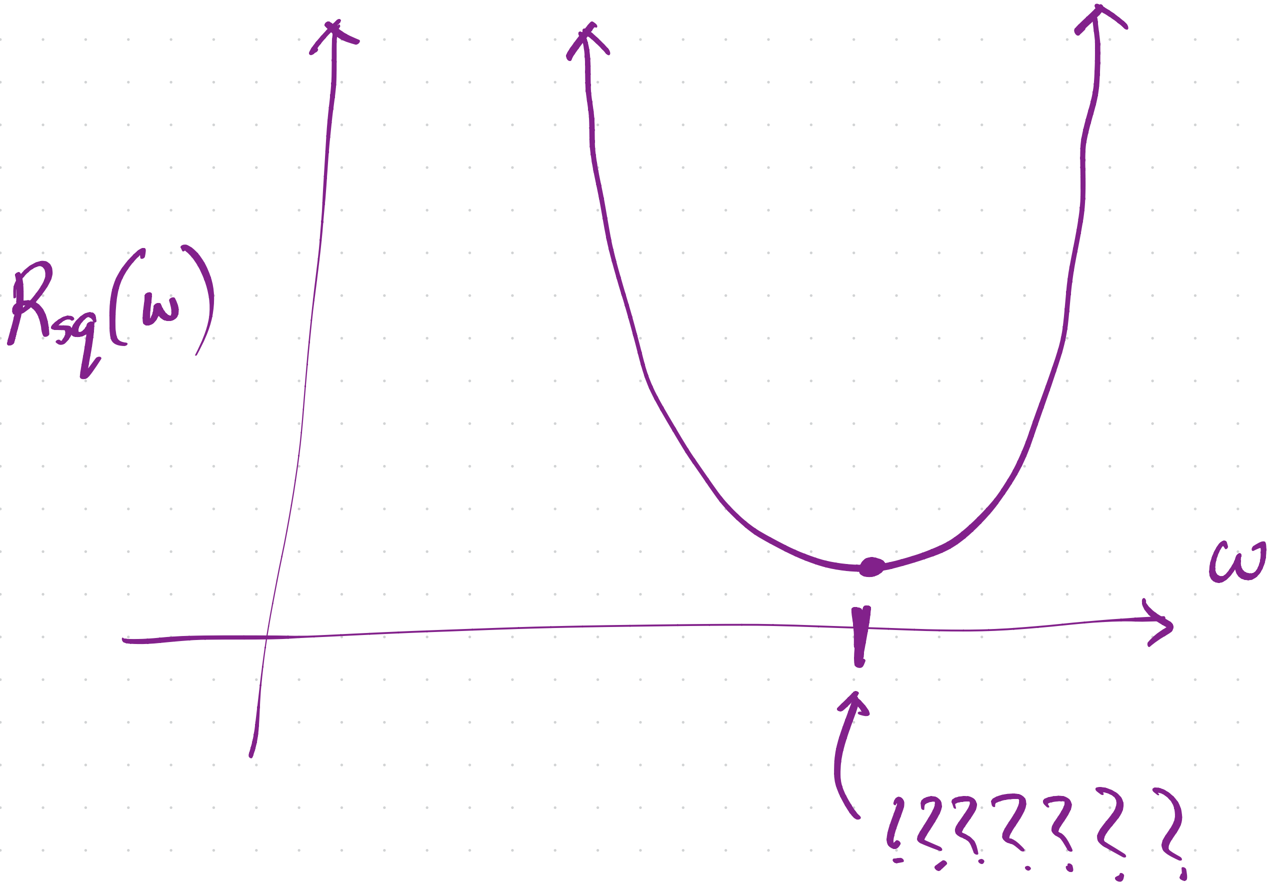
y_1 y_2 y_3 y_4 y_5

$$R_{sq}(w) = \frac{(72-w)^2 + (90-w)^2 + \cdots + (92-w)^2}{5}$$

L: loss for a single data point

R: average loss across whole dataset

empirical risk


$R_{sq}(w)$
w
??????

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

Goal: Find the best constant prediction, w

How? By minimizing average squared loss:

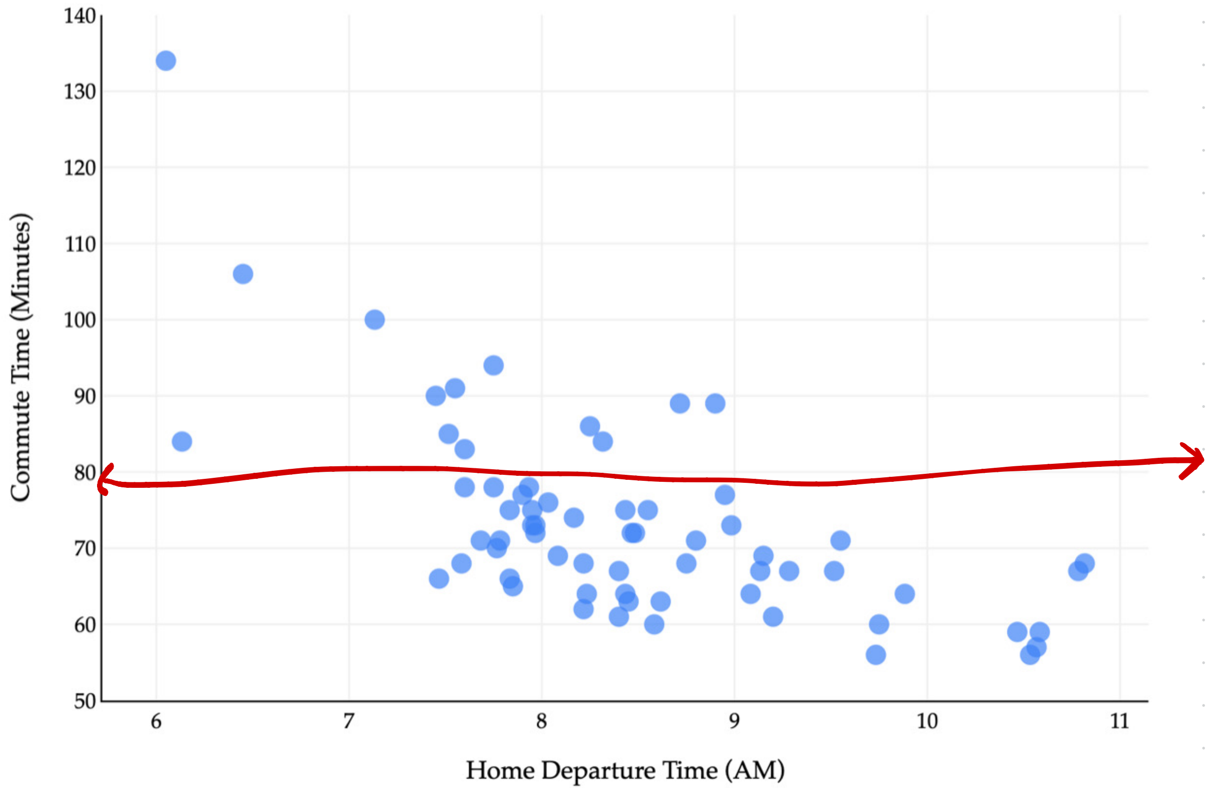
$$R_{sq}(w) = \frac{(y_1 - w)^2 + (y_2 - w)^2 + \cdots + (y_n - w)^2}{n}$$

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

need to minimize!

Terminology:

"average squared loss" = "mean squared error"



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

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

function of w only!

How to minimize function of one variable?

1. Take derivative with respect to w
2. Set equal to 0
3. Verify we found a minimum (rather than a maximum)

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

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

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

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

$$a(x) + b(x) + c(x) + \ldots$$

derivative: $\frac{da}{dx} + \frac{db}{dx} + \frac{dc}{dx}$

Aside:

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

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

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

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

set equal to 0 and solve for w

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

sum of errors is 0!

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

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

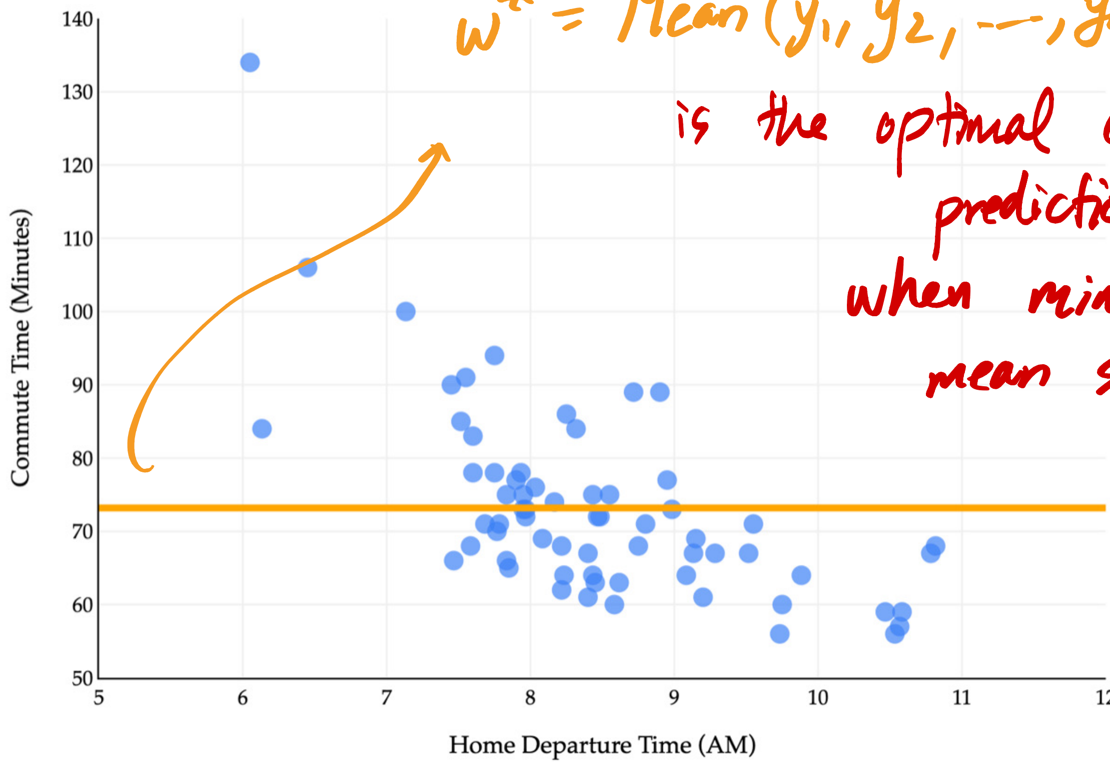
$w + w + \dots + w = nw$

Mean of y_i's is the best constant prediction!

w^* : optimal parameter
(best)

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

is the optimal constant prediction, when minimizing mean squared error



"Three-step modeling recipe" for making good predictions

① Choose a model

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

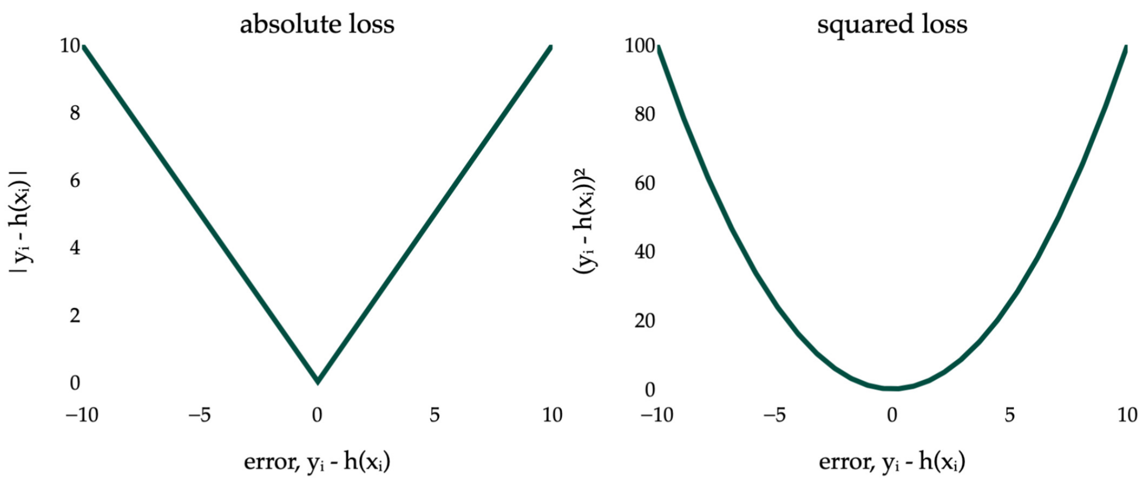
② Choose a loss function

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

③ Minimize average loss to find optimal parameters

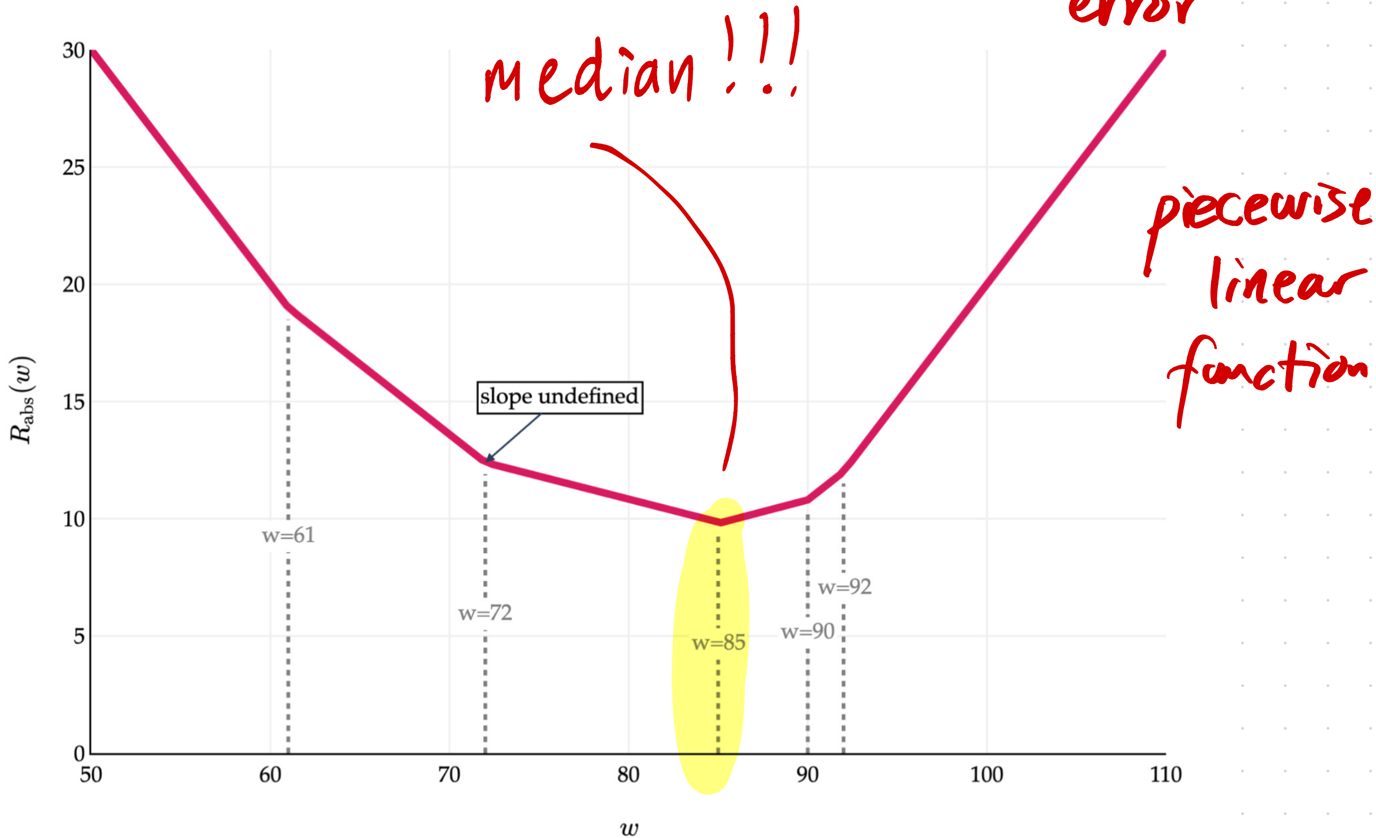
$$R_{sq}(w) = \frac{1}{n} \sum_{i=1}^{n} (y_i - w)^2 \rightarrow w^* = \text{Mean}(y_1, \ldots, y_n) = \bar{y}$$

<u>what if</u> we still use the constant model,
but use absolute loss instead?



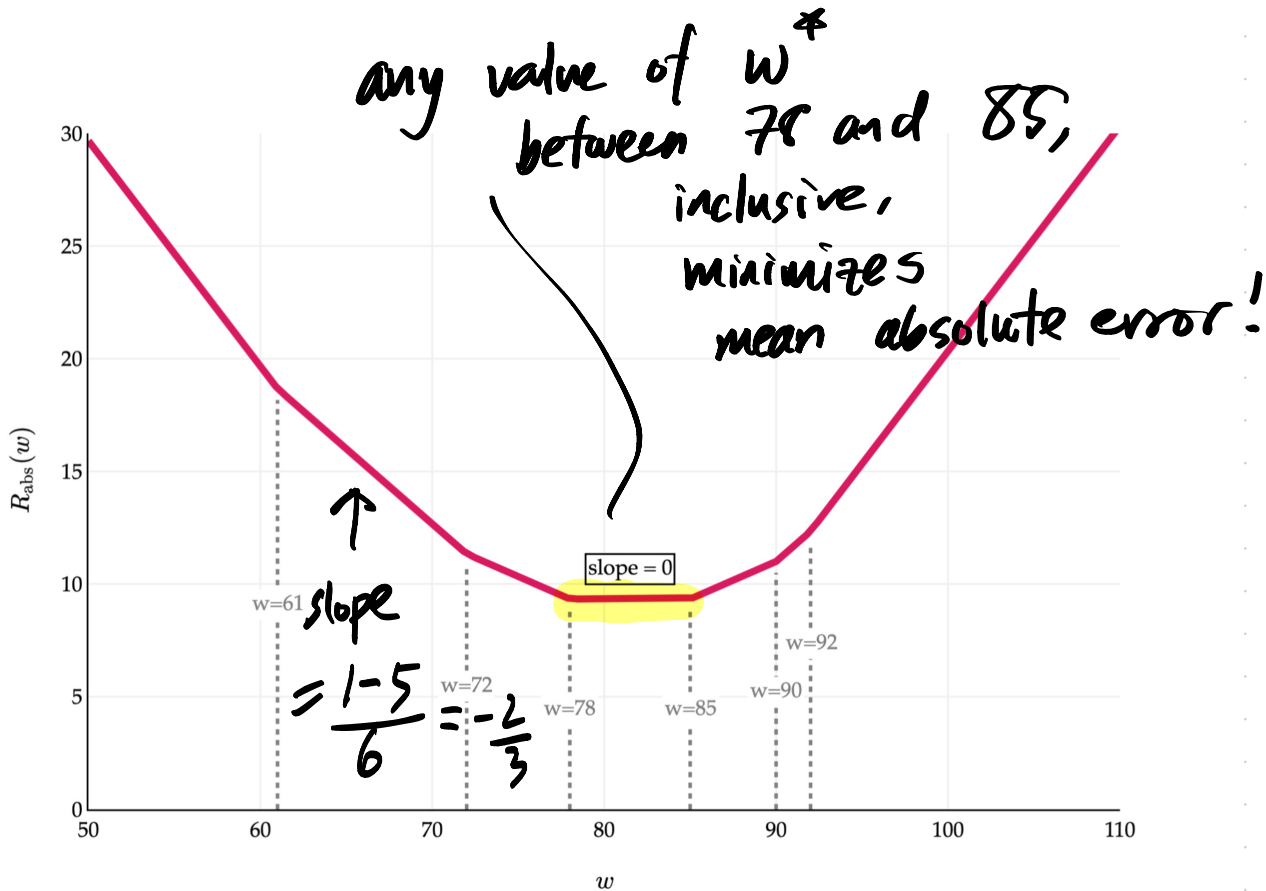
$$R_{\text{abs}}(w) = \frac{1}{5}(|72 - w| + |90 - w| + |61 - w| + |85 - w| + |92 - w|)$$

mean absolute error



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

And its graph is:



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