Lecture 2 – Minimizing Mean Absolute Error

DSC 40A, Spring 2023

Announcements

- Look at the readings linked on the course website!
- Discussion **tonight** at 7pm (A00), 8pm (B00) in FAH 1101.
  - Work on Groupwork 1 and submit it to Gradescope by tonight.
  - TAs and tutors will be there to help.
- Homework 1 is out and due Tuesday night.
- See Calendar on course website for office hours.
  - Plan to come to office hours at least once a week for help on homework.

Agenda

1. Recap from Lecture 1 – learning from data.
2. Minimizing mean absolute error.
3. Identifying another choice of error.

Recap from Lecture 1 – learning from data

Last time

- **Question:** How do we turn the problem of learning from data into a math problem?

- **Answer:** Through optimization.

A formula for the mean absolute error

- We have data:

$$90{,}000 \quad 94{,}000 \quad 96{,}000 \quad 120{,}000 \quad 160{,}000$$

- Suppose our prediction is h.
- The **mean absolute error** of our prediction is:

$$R(h) = \frac{1}{5}\Big(|90{,}000 - h| + |94{,}000 - h| + |96{,}000 - h| + |120{,}000 - h| + |160{,}000 - h|\Big)$$

Many possible predictions

- Last time, we considered four possible **hypotheses** for future salary, and computed the mean absolute error of each.
  - $h_1 = 150{,}000 \implies R(150{,}000) = 42{,}000$
  - $h_2 = 115{,}000 \implies R(115{,}000) = 23{,}000$
  - $h_3 = \text{mean} = 112{,}000 \implies R(112{,}000) = 22{,}400$
  - $h_4 = \text{median} = 96{,}000 \implies R(96{,}000) = 19{,}200$
- Of these four options, the median has the lowest MAE. But is it the **best possible prediction overall**?

A *general* formula for the mean absolute error

- Suppose we collect n salaries, $y_1, y_2, \ldots, y_n$.
- The mean absolute error of the prediction h is:

$$\begin{aligned} R(h) &= \frac{1}{n}\left(|h - y_1| + |h - y_2| + \ldots + |h - y_n|\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} |h - y_i| \end{aligned}$$

The best prediction

- We want the best prediction, h^*.
- The smaller $R(h)$, the better h.
- Goal: find h that minimizes $R(h)$.

Discussion Question

Can we use calculus to minimize R?

Minimizing mean absolute error

Minimizing with calculus

- Calculus: take derivative with respect to h, set equal to zero, solve.

$$R(h) = \frac{1}{n}\left(\sum_{i=1}^{n} |h - y_i|\right)$$

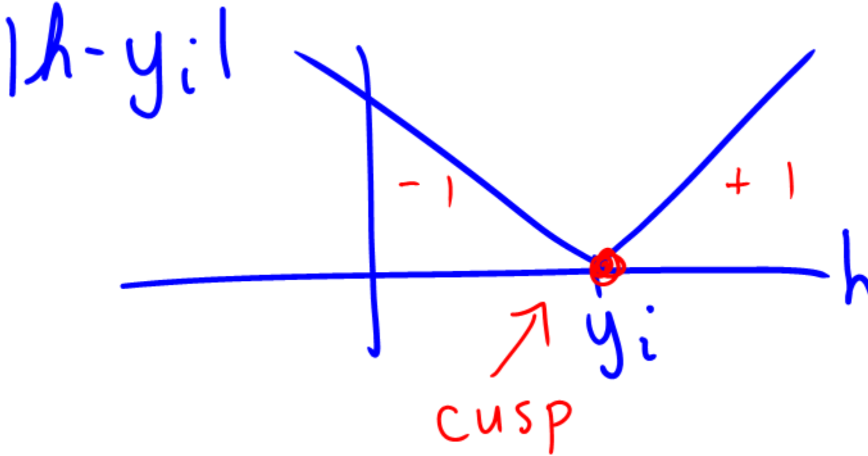
$$R'(h) = \frac{1}{n}\left(\sum_{i=1}^{n} \frac{d}{dh}\underbrace{(|h - y_i|)}_{\text{not differentiable}}\right)$$

$$f(x) = c \cdot g(x)$$
$$f'(x) = c \cdot g'(x)$$

$$f(x) = g(x) + h(x)$$
$$f'(x) = g'(x) + h'(x)$$

Minimizing with calculus

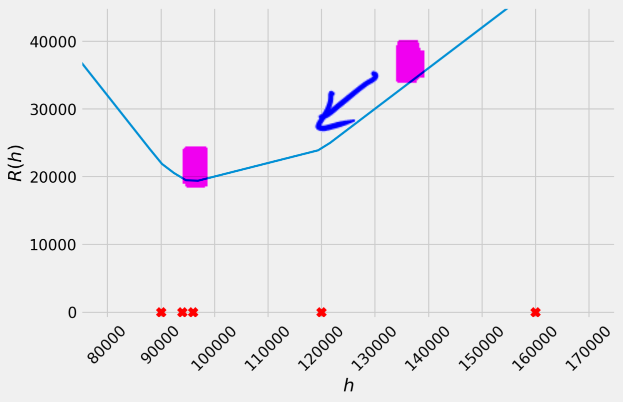
- Calculus: take derivative with respect to *h*, set equal to zero, solve.



Uh oh...

- R is **not differentiable**.
- We can't use calculus to minimize it.
- Let's try plotting $R(h)$ instead.

Plotting the mean absolute error



1) continuous
2) made up of line segments

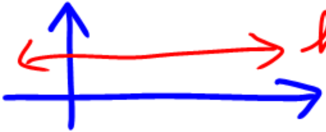
$\rightarrow$ neg. to nonneg.

Discussion Question

A local minimum occurs when the slope goes from ______________. Select all that apply.

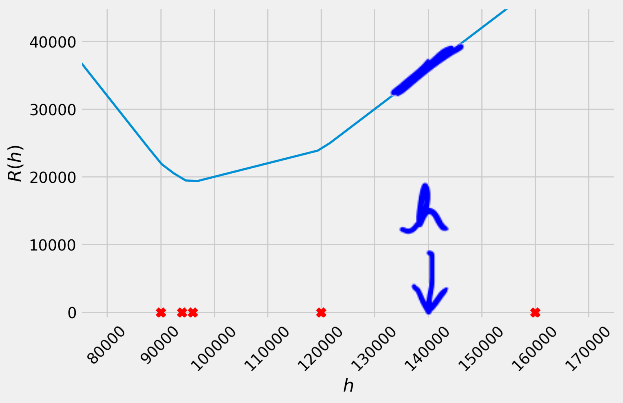
A) positive to negative
B) negative to positive
C) positive to zero.
D) negative to zero.

ex.) $f(x)=5$



has min everywhere!

Goal



- Find where slope of R goes from negative to non-negative.
- Want a formula for the slope of R at h.

Sums of linear functions

- Let

$$f_1(x) = 3x + 7 \qquad f_2(x) = 5x - 4 \qquad f_3(x) = -2x - 8$$

- What is the slope of $f(x) = f_1(x) + f_2(x) + f_3(x)$?

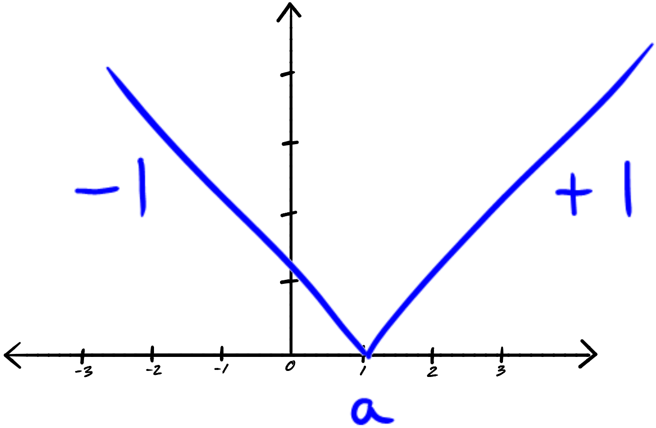
$$3x + 7 + 5x - 4 + -2x + 8$$

$$3x + 5x - 2x + c$$

$$6x + c$$

Absolute value functions

Recall, $f(x) = |x - a|$ is an absolute value function centered at $x = a$.



Sums of absolute values

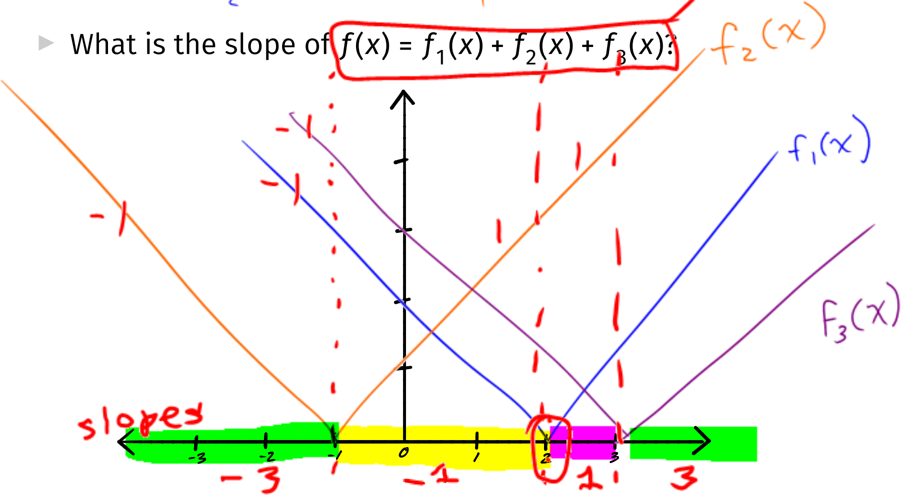
minimized at $x = 2$

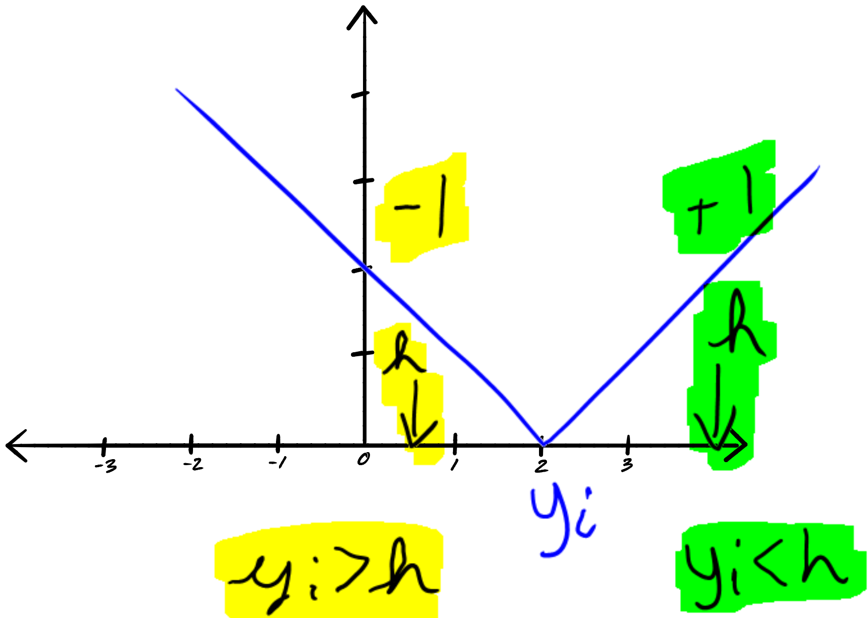
- Let

$$f_1(x) = |x - 2| \qquad f_2(x) = |x + 1| \qquad f_3(x) = |x - 3|$$

$$\qquad\quad 2 \qquad\qquad\qquad\qquad -1$$

- What is the slope of $\boxed{f(x) = f_1(x) + f_2(x) + f_3(x)}$?





-1
+1
h
h
-3
-2
-1
0
1
2
3
y_i
$y_i > h$
$y_i < h$

$-(-2) = |-2|$

$a < b$ or
$a > b$ or
$a = b$

The slope of the mean absolute error

$R(h)$ is a sum of absolute value functions (times $\frac{1}{n}$):

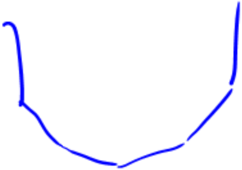
$$R(h) = \frac{1}{n}\left(|h - y_1| + |h - y_2| + \ldots + |h - y_n|\right)$$

$$R(h) = \frac{1}{n}\left(\sum_{y_i < h} \underbrace{|h - y_i|} + \sum_{y_i > h} |h - y_i| + \sum_{y_i = h} |h - y_i|\right)$$

$$= \frac{1}{n}\left(\sum_{y_i < h} (h - y_i) + \sum_{y_i > h} -(h - y_i) + 0\right)$$

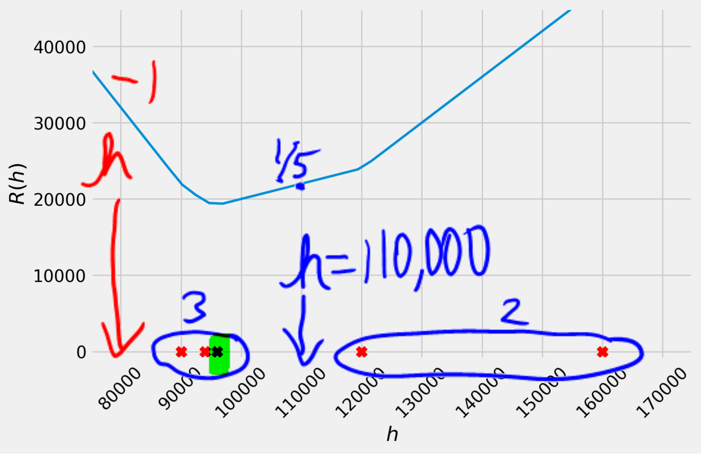
$$\text{slope of } R(h) = \frac{1}{n}\left(1 * \# y_i < h + -1 * \# y_i > h\right)$$

The slope of the mean absolute error



The slope of R at h is: $\frac{1}{5}(0 - 5)$

$$\frac{1}{n} \cdot \left[\underbrace{(\#\text{ of } y_i\text{'s} < h)} - (\#\text{ of } y_i\text{'s} > h) \right]$$



Where the slope's sign changes

The slope of R at h is:

$$\frac{1}{n} \cdot \left[(\# \text{ of } y_i\text{'s} < h) - (\# \text{ of } y_i\text{'s} > h) \right]$$

Discussion Question

Suppose that n is odd. At what value of h does the slope of R go from negative to non-negative?

A) h = mean of $y_1, \ldots, y_n$
B) h = median of $y_1, \ldots, y_n$
C) h = mode of $y_1, \ldots, y_n$

The median minimizes mean absolute error, when n is odd

- Our problem was: find h^* which minimizes the mean absolute error, $R(h) = \frac{1}{n}\sum_{i=1}^{n} |y_i - h|$.
- We just determined that when n is odd, the answer is $\text{Median}(y_1, \ldots, y_n)$. This is because the median has an equal number of points to the left of it and to the right of it.
- But wait — what if n is **even**?

slope of $R(h) = \frac{1}{n}(\# y_i < h - \# y_i > h)$

Discussion Question

Consider again our example dataset of 5 salaries.

90,000 94,000 96,000 120,000 160,000

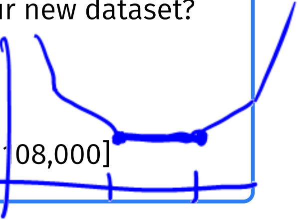
$-4/6$ $-2/6$ 0

Suppose we collect a 6th salary, so that our data is now

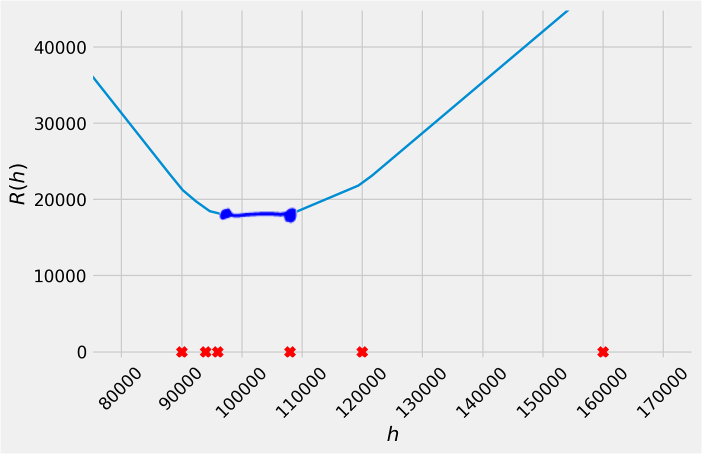
90,000 94,000 96,000 108,000 120,000 160,000

Which of the following correctly describes the h^* that minimizes mean absolute error for our new dataset?

A) 96,000 only
B) 108,000 only
C) 102,000 only
D) Any value in the interval [96,000, 108,000]



Plotting the mean absolute error, with an even number of data points



- What do you notice?

The median minimizes mean absolute error

- Our problem was: find h^* which minimizes the mean absolute error, $R(h) = \frac{1}{n} \sum_{i=1}^{n} |y_i - h|$.

Answer: the median!

- **Regardless of if n is odd or even**, the answer is $h^* = \text{Median}(y_1, \dots, y_n)$. The **best prediction**, in terms of mean absolute error, is the **median**.
  - When n is odd, this answer is unique.
  - When n is even, any number between the middle two data points also minimizes mean absolute error.
  - We define the median of an even number of data points to be the mean of the middle two data points.

Identifying another type of error

Two things we don't like

1. **Minimizing** the mean absolute error wasn't so easy.

2. Actually **computing** the median isn't so easy, either.

- **Question**: Is there another way to measure the quality of a prediction that avoids these problems?

The mean absolute error is not differentiable

- We can't compute $\frac{d}{dh}|y_i - h|$.

- Remember: $|y_i - h|$ measures how far h is from y_i.

- Is there something besides $|y_i - h|$ which:
  1. Measures how far h is from y_i, *and*
  2. is **differentiable**?

keep intuition

The mean absolute error is not differentiable

- We can't compute $\frac{d}{dh}|y_i - h|$.
- Remember: $|y_i - h|$ measures how far h is from y_i.
- Is there something besides $|y_i - h|$ which:
  1. Measures how far h is from y_i, *and*
  2. is **differentiable**?

→ intuitive

→ diff

Discussion Question

Which of these would work?

a) $e^{|y_i - h|}$ ✓ ✗

b) $|y_i - h|^2$ ✓ ✓

because $(|y_i - h|)^2 = (y_i - h)^2$

c) $|y_i - h|^3$ ✓ ✗

d) $\cos(y_i - h)$ ✗ ✗

Why?

Summary

Summary

- Our first problem was: find h^* which minimizes the mean absolute error, $R(h) = \frac{1}{n}\sum_{i=1}^{n} |y_i - h|$.
  - The answer is: $\text{Median}(y_1, \dots, y_n)$.
  - The **best prediction**, in terms of mean absolute error, is the **median**.
- We then started to consider another type of error that is differentiable and hence is easier to minimize.
- **Next time:** We will find the value of h^* that minimizes this other error, and see how it compares to the median.