

## Lecture 2 – Minimizing Mean Absolute Error



DSC 40A, Fall 2022 @ UC San Diego

Mahdi Soleymani, with help from [many others](#)

# Announcements

- ▶ Look at the readings linked on the course website!
- ▶ Homework 1 will be out Friday September 30th and is due on **Friday, 10/7 at 2:00pm.**
- ▶ Groupwork 1 will be out on Thursday September 29th and is due on **11:59pm Monday, October 3rd.**
- ▶ Come to discussion **next Monday** to work on groupwork.
- ▶ See Calendar on course website for office hours. Location: SDSC.
- ▶ Access to Campuswire, Gradescope (3JDXW7), and, **Podcasts.**

# Agenda

1. Recap from Lecture 1 – learning from data.
2. Minimizing mean absolute error.

## **Recap from Lecture 1 – learning from data and optimization**

## Mean absolute error (MAE)

- ▶ Mean absolute error on data:

$$h_1 : 42,000 \quad h_2 : 23,000$$

- ▶ Conclusion:  $h_2$  is the better prediction.
- ▶ In general: pick prediction with the smaller mean absolute error.

## We are making an assumption...

- ▶ We're assuming that future salaries will look like present salaries.
- ▶ That a prediction that was good in the past will be good in the future.

### Discussion Question

Is this a good assumption?

## Which is better: the mean or median?

- ▶ Recall:

mean = 112,000      median = 96,000

- ▶ We can calculate the mean absolute error of each:

mean : 22,400      median : 19,200

- ▶ The median is the best prediction so far!
- ▶ But is there an even better prediction?

## Finding the best prediction

- ▶ Any (non-negative) number is a valid prediction.
- ▶ Goal: out of all predictions, find the prediction  $h^*$  with the smallest mean absolute error.
- ▶ This is an **optimization problem**.



## A formula for the mean absolute error

- ▶ We have data:

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

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

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

## A formula for the mean absolute error

- ▶ We have a function for computing the mean absolute error of **any** possible prediction.

$$\begin{aligned}R(150,000) &= \frac{1}{5} \left( |90,000 - 150,000| + |94,000 - 150,000| \right. \\ &\quad + |96,000 - 150,000| + |120,000 - 150,000| \\ &\quad \left. + |160,000 - 150,000| \right) \\ &= 42,000\end{aligned}$$

## A formula for the mean absolute error

- ▶ We have a function for computing the mean absolute error of **any** possible prediction.

$$\begin{aligned}R(\mathbf{115,000}) &= \frac{1}{5} \left( |90,000 - \mathbf{115,000}| + |94,000 - \mathbf{115,000}| \right. \\ &\quad + |96,000 - \mathbf{115,000}| + |120,000 - \mathbf{115,000}| \\ &\quad \left. + |160,000 - \mathbf{115,000}| \right) \\ &= \mathbf{23,000}\end{aligned}$$

## A formula for the mean absolute error

- ▶ We have a function for computing the mean absolute error of **any** possible prediction.

$$\begin{aligned}R(\pi) &= \frac{1}{5} \left( |90,000 - \pi| + |94,000 - \pi| \right. \\ &\quad + |96,000 - \pi| + |120,000 - \pi| \\ &\quad \left. + |160,000 - \pi| \right) \\ &= \mathbf{111,996.8584\dots}\end{aligned}$$

### Discussion Question

Without doing any calculations, which is correct?

- A.  $R(50) < R(100)$
- B.  $R(50) = R(100)$
- C.  $R(50) > R(100)$

## A *general* formula for the mean absolute error

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

- ▶ Or, using **summation notation**:
-

## A formula for the mean absolute error

- ▶ We have data:

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

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

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

## 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**?

## 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**

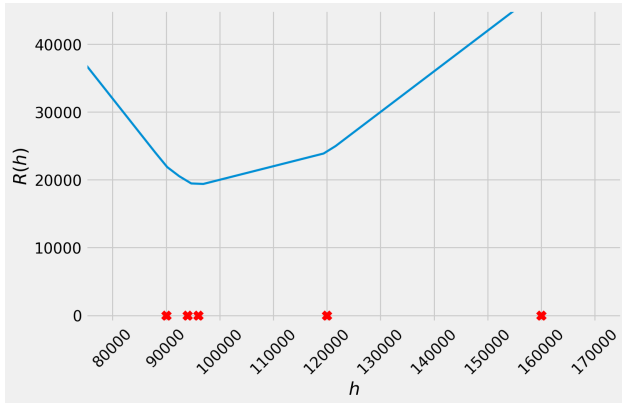
## Minimizing with calculus

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

## 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

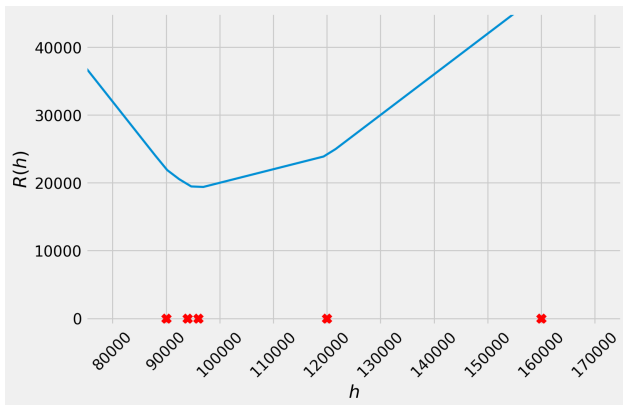


## Discussion Question

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.

# 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

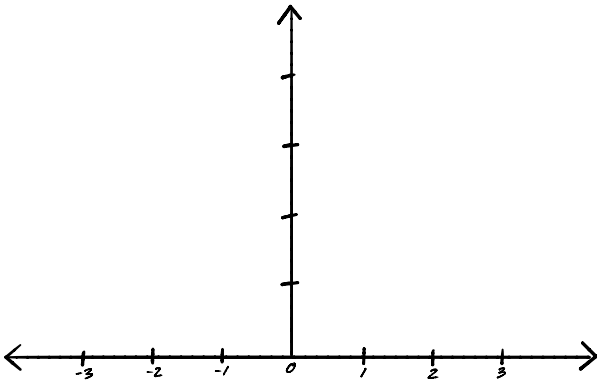
- ▶ Let

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

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

## Absolute value functions

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



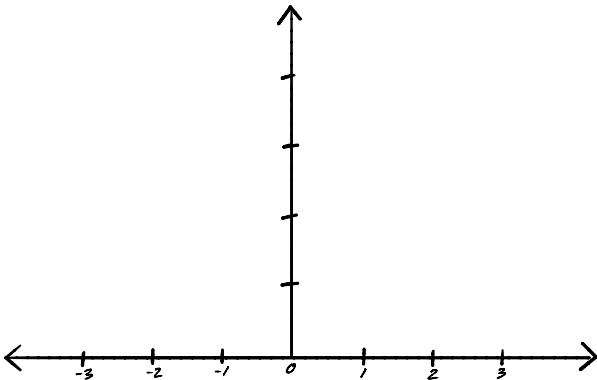


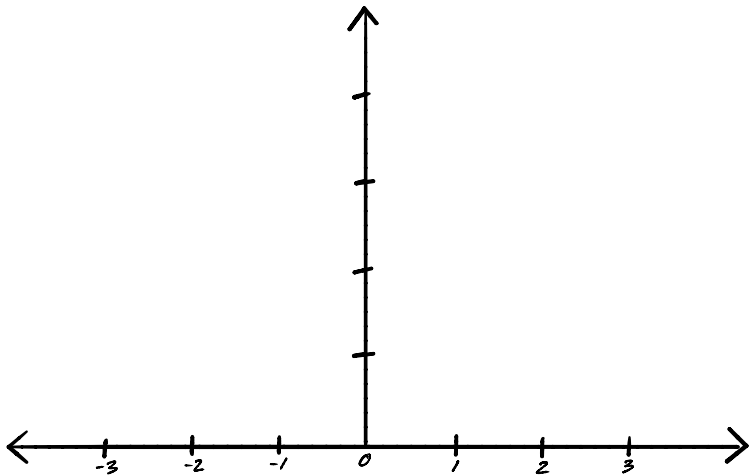
## Sums of absolute values

- ▶ Let

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

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





## The slope of the mean absolute error

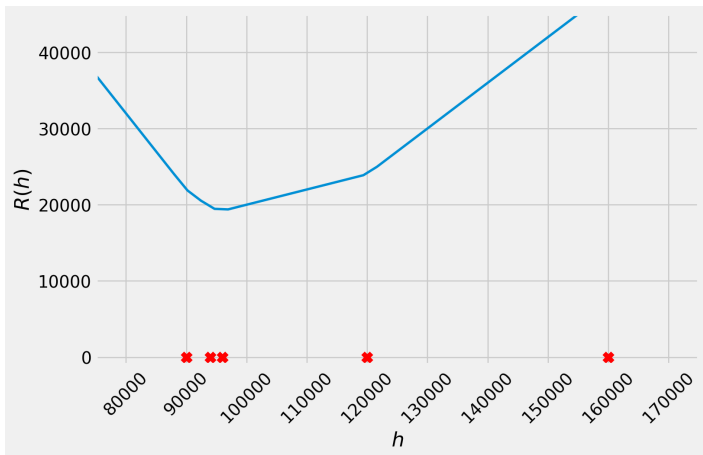
$R(h)$  is a sum of absolute value functions (times  $\frac{1}{n}$ ):

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

# The slope of the mean absolute error

The slope of  $R$  at  $h$  is:

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



## Where the slope's sign changes

The slope of  $R$  at  $h$  is:

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

### Discussion Question

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

- A)  $h = \text{mean of } y_1, \dots, y_n$
- B)  $h = \text{median of } y_1, \dots, y_n$
- C)  $h = \text{mode of } y_1, \dots, y_n$

## 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  $\text{Median}(y_1, \dots, y_n)$ . 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**?

## Discussion Question

Consider again our example dataset of 5 salaries.

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

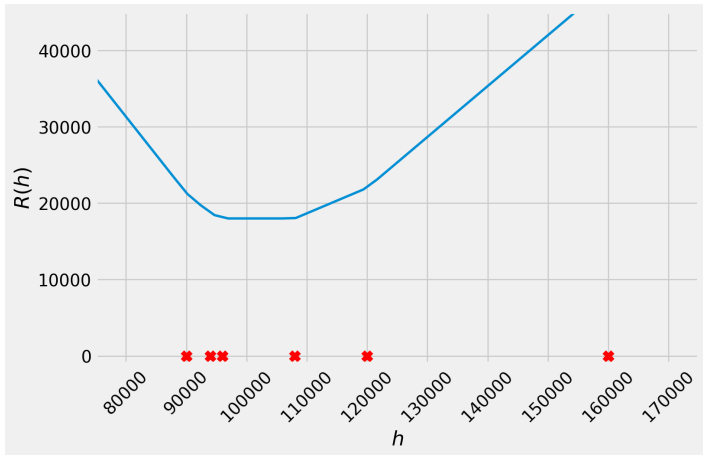
Suppose we collect a 6th salary, so that our data is now

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 between 96,000 and 108,000, inclusive

# 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,  $R(h) = \frac{1}{n} \sum_{i=1}^n |y_i - h|$ .

- ▶ **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**?

## 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**?

### Discussion Question

Which of these would work?

a)  $e^{|y_i - h|}$

b)  $|y_i - h|^2$

c)  $|y_i - h|^3$

d)  $\cos(y_i - h)$