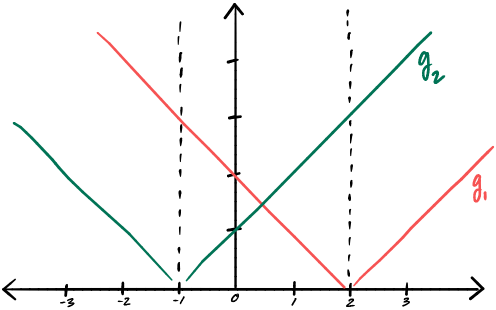


DSC 40A

Lecture 02

Learning via Optimization, pt II

Announcements

- Extension students: join Gradescope/Campuswire using codes found on `www.dsc40a.com`

- Need iClicker starting next week for tokens.

Last Time

How do we turn the problem of learning into a math problem?

Last Time

- What will be your future salary?
- Collect data:

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

- Could use the **mean** or the **median** as a prediction.
- But why?
- What is the best prediction?

Last Time: The Mean Error of a Prediction

- Suppose we predicted a future salary of $h_1 = 150{,}000$ *before* collecting data.

salary	error of h_1
90,000	60,000
94,000	56,000
96,000	54,000
120,000	30,000
160,000	10,000
	total error: 210,000
	mean error: 42,000

- A good prediction is one with small mean error.

Last Time: The Best Prediction

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

Today

We’ve turned learning into an **optimization problem**.
How do we solve it?

A Formula for the Mean 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 error** of our prediction is:

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

A Formula for the Mean Error

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

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

A Formula for the Mean Error

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

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

A Formula for the Mean Error

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

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

A *General* Formula for the Mean Error

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

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

- Or, using **summation notation**:

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

$1, 2, \ldots, 16$

The Best Prediction

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

$$\frac{d}{dx}\left[f(x) + g(x)\right] = \frac{df}{dx} + \frac{dg}{dx}$$

Discussion Question

Can we use calculus to minimize R?

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

$$= \frac{1}{n}\sum_{i=1}^{n} |y_i - h|$$

Minimizing with Calculus

wrt h

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

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

$$\frac{dR}{dh} = \frac{d}{dh}\left[\frac{1}{n} \sum_{i=1}^{n} |y_i - h|\right]$$

$$= \frac{1}{n}\left[\frac{d}{dh} \sum_{i=1}^{n} |y_i - h|\right]$$

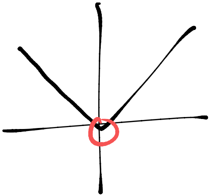
$$= \frac{1}{n} \sum_{i=1}^{n} \frac{d}{dh} |y_i - h|$$

Minimizing with Calculus

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

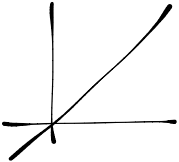
$f(x) = |x|$

$\frac{df}{dx} = ?$



$g(x) = x$

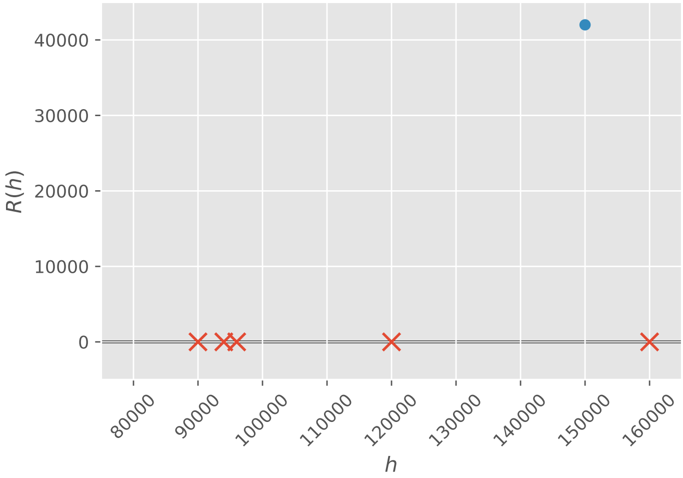
$\frac{dg}{dx} = 1$



Uh oh

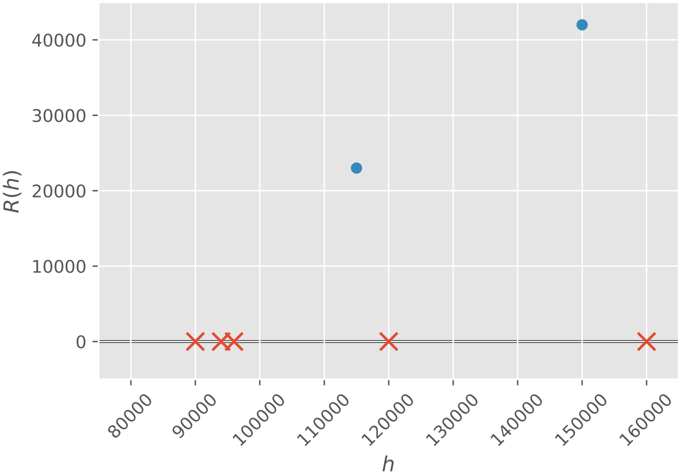
- R is **not differentiable**.
- We can't use calculus to minimize it.
- Let's try plotting $R(h)$

Plotting the Mean Error



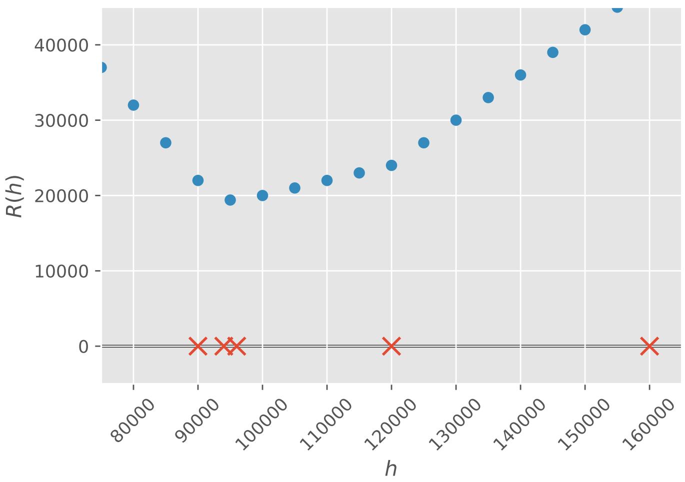
Recall: $R(150{,}000) = 42{,}000$

Plotting the Mean Error

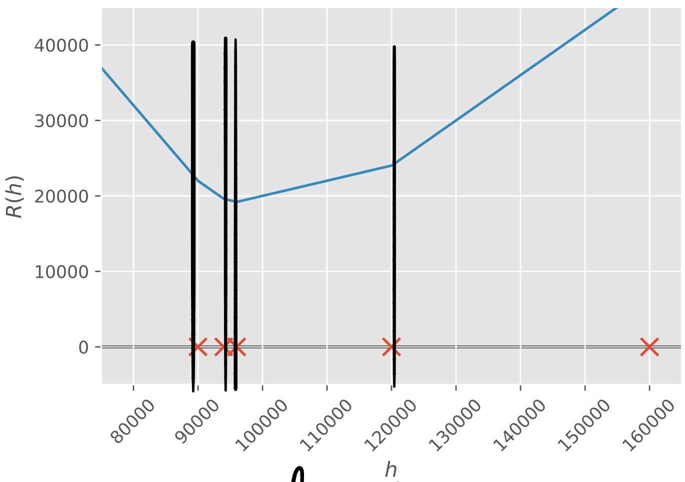


Recall: $R(115{,}000) = 23{,}000$

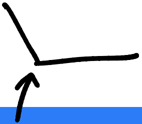
Plotting the Mean Error

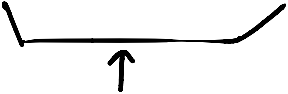


Plotting the Mean Error



"piecewise linear"

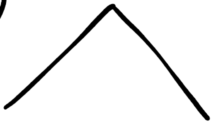
d) 

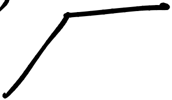
e) 

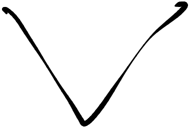
Discussion Question

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

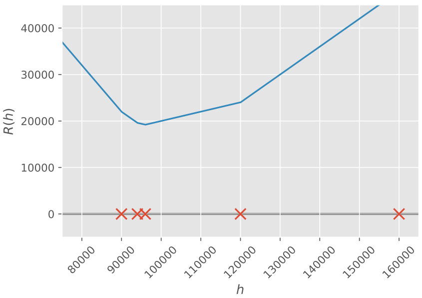
~~A)~~ positive to negative
B) negative to positive
C) positive to zero.
D) negative to zero.
E) zero to zero.

a) 

c) 

b) 

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 $f_1(x) = 3x + 2$
- Let $f_2(x) = 5x + 1$
- What is the slope of $f(x) = f_1(x) + f_2(x)$?

$$f(x) = (3x+2) + (5x+1)$$

$$= 8x + 3$$

Sums of Absolute Values

$$g_2(x) = \begin{cases} x-1 & x > -1 \\ 1-x & x < -1 \end{cases}$$

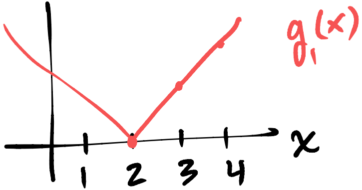
- Let

$$g_1(x) = |x - 2| \qquad g_2(x) = |x + 1|$$

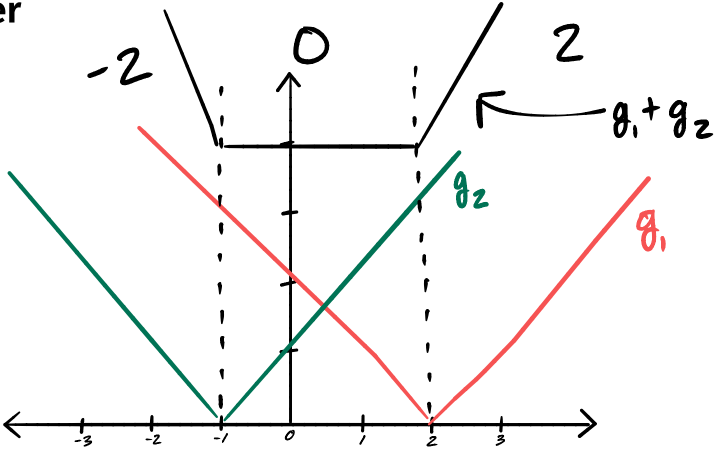
- Let $g(x) = g_1(x) + g_2(x)$.

Discussion Question

What is the slope of g at $x = 1$?



Answer



$g_1(x) = |x-2|$

$g_2(x) = |x+1|$

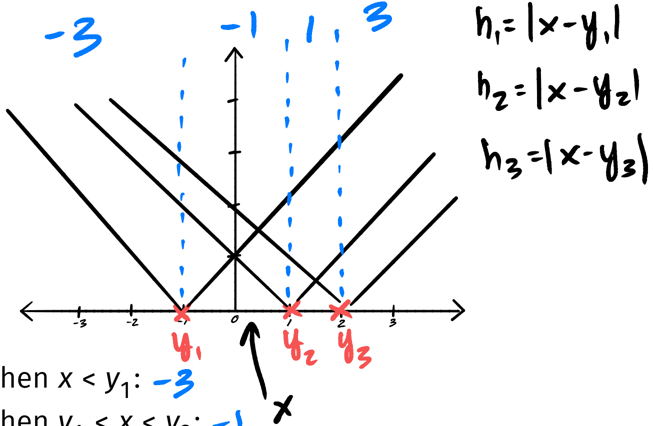
Sums of More Absolute Values

- Let $y_1 < y_2 < y_3$

$$h_1(x) = |x - y_1| \qquad h_2(x) = |x - y_2| \qquad h_3(x) = |x - y_3|$$

- Let $h(x) = h_1(x) + h_2(x) + h_3(x)$.
- The slope changes at y_1, y_2, y_3.

Sums of More Absolute Values



- Slope when $x < y_1$: -3
- Slope when $y_1 < x < y_2$: -1
- Slope when $y_2 < x < y_3$: 1
- Slope $x > y_3$: 3

Slope at x = (# of y_i's $\underline{<}$ x) - (# of y_i's $\underline{>}$ x)

The Slope of Error Function

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

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

- So:

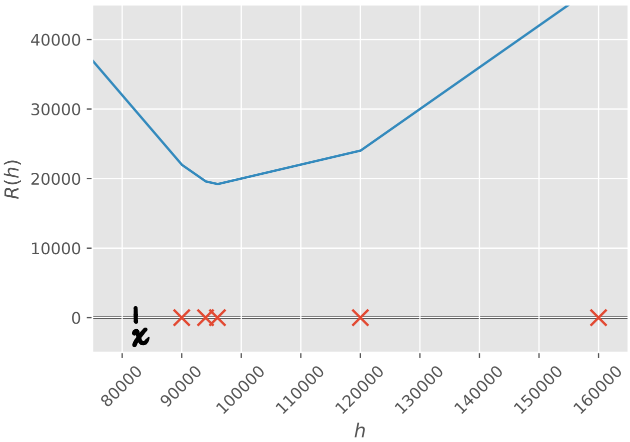
Slope at h = $\frac{1}{n} \cdot \left[(\# \text{ of } y_i's \ \underline{<} \ h) - (\# \text{ of } y_i\text{'s} \ \underline{>} \ h)\right]$

Discussion Question

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

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

Where the Slope’s Sign Changes



The Median Minimizes the Mean Error

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

- The answer is: $\text{Median}(y_1, \ldots, y_n)$.

- The **best prediction**[1] is the **median**.

[1]in terms of mean error

Status Update

- Last time, we turned predicting salary into a math problem: minimize the mean error.

- Today: we solved it. The **median** minimizes the mean error.

What's Left?

- We did all this because $R(h)$ isn't differentiable.
- What if we tried to minimize a *different* measure of error that *is* differentiable?