DSC 40A - Extra Practice for Final Part 1

Spring 2023

Problem 1. Sloped Mean

Suppose you have a data set $y_1, y_2, \ldots, y_n$ with at least three values, $n \geq 3$, and the values are arranged such that $y_1 \leq y_2 \leq \cdots \leq y_n$.

We know from class that the mean of the data minimizes mean squared error,

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

Define a new function that weights larger data points less heavily:

$$MSE = \sum_{i=1}^{n} (h-y_i)^2$$

$$S(h) = \left(\sum_{i=1}^{n-2} (h - y_i)^2 \right) + 0.5 \cdot (h - y_{n-1})^2 + 0.1 \cdot (h - y_n)^2.$$

a) What value of h minimizes $S(h)$? We'll call the value of h that minimizes $S(h)$ the **sloped mean**, since the coefficients of the data values decrease for larger data.

Solution:

$$h = \frac{\left(\sum_{i=1}^{n-2} y_i \right) + 0.5 \cdot y_{n-1} + 0.1 \cdot y_n}{n - 1.4}$$

We can prove this using calculus.

$$\sum_{i=1}^{n-2} (h - y_i) \neq (n-2)(h-y_i)$$

$$S'(h) = \left(\sum_{i=1}^{n-2} 2 \cdot (h - y_i) \right) + 2 \cdot 0.5 \cdot (h - y_{n-1}) + 2 \cdot 0.1 \cdot (h - y_n)$$

$$0 = 2 \cdot \left(\left(\sum_{i=1}^{n-2} (h - y_i) \right) + 0.5 \cdot (h - y_{n-1}) + 0.1 \cdot (h - y_n) \right)$$

$$\sum_{i=1}^{n-2} h - \sum_{i=1}^{n-2} y_i$$

isolate h $\rightarrow$

$$\boxed{0 = \left(\sum_{i=1}^{n-2} (h - y_i) \right) + 0.5 \cdot (h - y_{n-1}) + 0.1 \cdot (h - y_n)}$$

$$0 = (n-2) \cdot h - \left(\sum_{i=1}^{n-2} y_i \right) + 0.5 \cdot h - 0.5 \cdot y_{n-1} + 0.1 \cdot h - 0.1 \cdot y_n$$

$$\underbrace{\sum_{i=1}^{n-2} h}_{(n-2)h}$$

$$(n-2) \cdot h + 0.5 \cdot h + 0.1 \cdot h = \left(\sum_{i=1}^{n-2} y_i \right) + 0.5 \cdot y_{n-1} + 0.1 \cdot y_n$$

$$h \cdot (n - 1.4) = \left(\sum_{i=1}^{n-2} y_i \right) + 0.5 \cdot y_{n-1} + 0.1 \cdot y_n$$

$$h = \frac{\left(\sum_{i=1}^{n-2} y_i \right) + 0.5 \cdot y_{n-1} + 0.1 \cdot y_n}{n - 1.4}$$

We know that this critical point corresponds to a minimum because $S(h)$ is a quadratic with a positive leading coefficient, so it's an upward-facing parabola.

b) Which do you think is a better hypothesis, the mean or the sloped mean? Is your answer always the same, or does it depend on some property of the data set? Give an example of when you might prefer to use the sloped mean, and when you might prefer the (regular) mean.

Problem 2. Which is bigger? By how much?

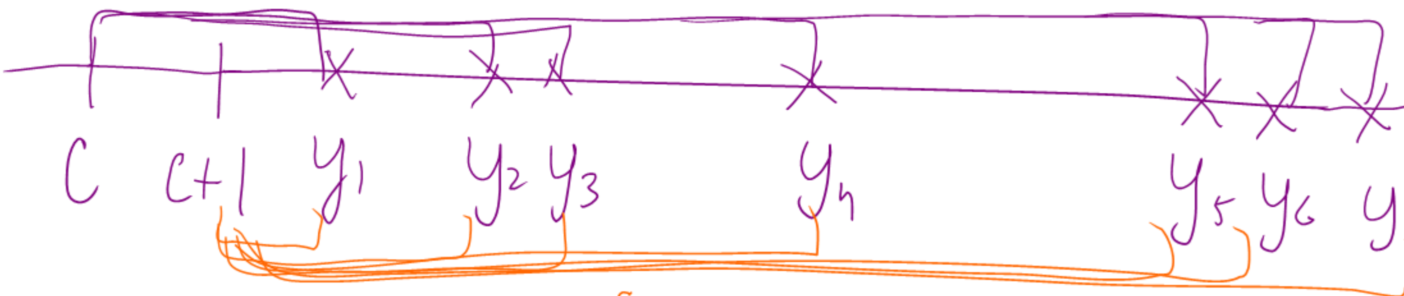
Given a data set $y_1 \leq y_2 \leq \cdots \leq y_n$, define the following empirical risk functions:

MAE $$R_{\text{abs}}(h) = \frac{1}{n}\sum_{i=1}^{n}|y_i - h| \qquad\qquad R_{\text{sq}}(h) = \frac{1}{n}\sum_{i=1}^{n}(y_i - h)^2$$

Parts (a), (b), and (c) below concern R_{abs}. Parts (d) and (e) concern R_{sq}.

a) For an arbitrary c with $c < c+1 < y_1$, how does $R_{\text{abs}}(c)$ compare to $R_{\text{abs}}(c+1)$? Can you determine which is bigger, and by how much?

one less than $R_{abs}(c)$

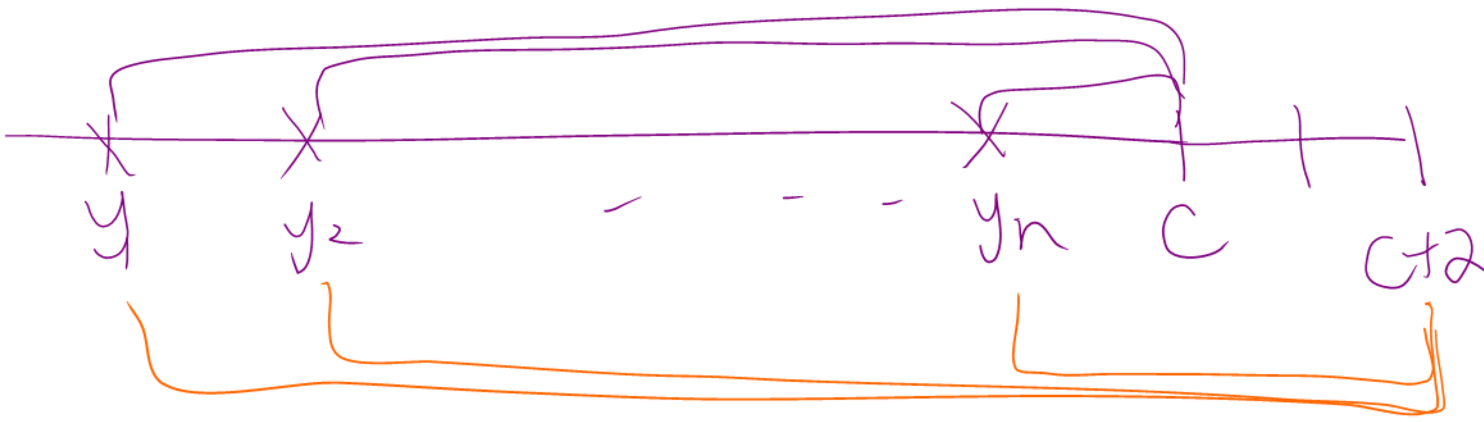


1 unit closer for every data point when you change prediction to c+1

$\Rightarrow$ 1 unit closer on avg

$$R_{abs}(c+1) = R_{abs}(c) - 1$$

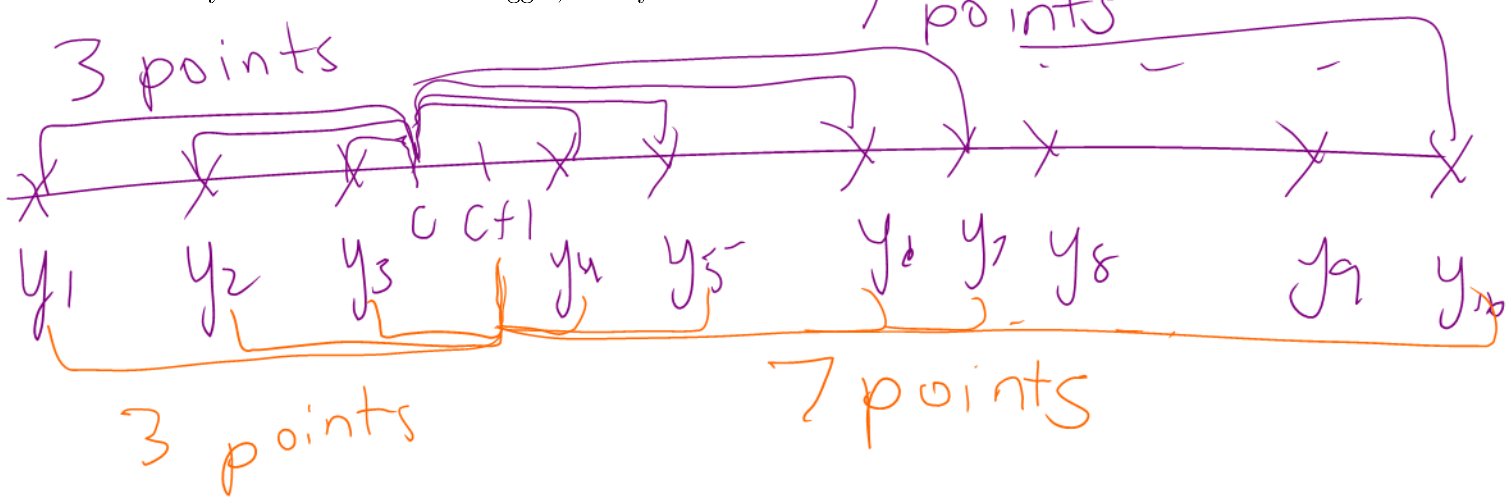
b) For an arbitrary c with $y_n < c < c + 2$, how does $R_{\text{abs}}(c)$ compare to $R_{\text{abs}}(c+2)$? Can you determine which is bigger, and by how much?



every line 2 units longer!

$$R_{abs}(c+2) = R_{abs}(c) + 2$$

c) Suppose $n = 10$. For an arbitrary c with $y_3 < c < c{+}1 < y_4$, how does $R_{\text{abs}}(c)$ compare to $R_{\text{abs}}(c{+}1)$? Can you determine which is bigger, and by how much?



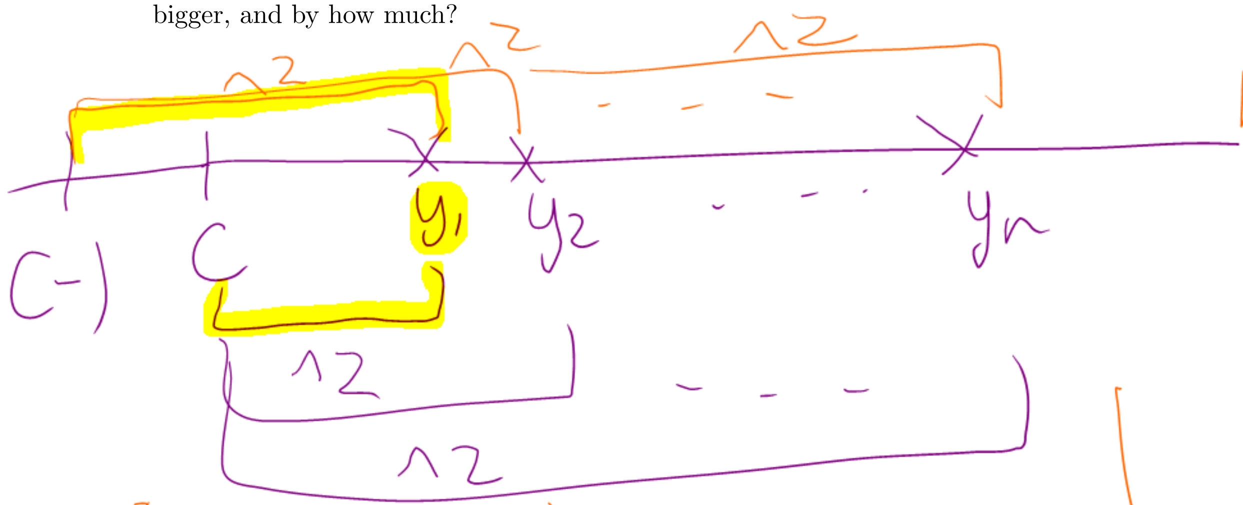
total distance to all points from c+1 = total distance to all points from c +3 − 7

divide by 10

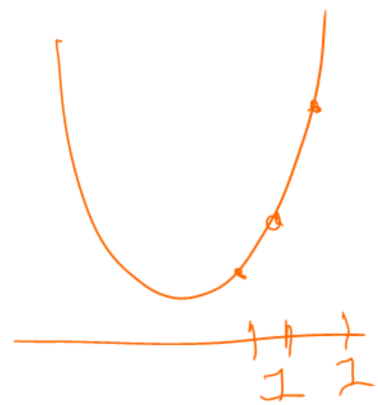
$R_{abs}(c+1) = R_{abs}(c) - 0.4$

$R_{sq}(c-1) > R_{sq}(c)$ but not sure by how much

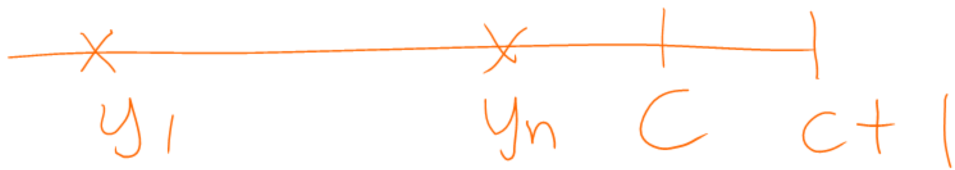
d) For an arbitrary c with $c < y_1$, how does $R_{\text{sq}}(c)$ compare to $R_{\text{sq}}(c-1)$? Can you determine which is bigger, and by how much?



dist c to y_1	sq dist c to y_1	dist c-1 to y_1	sq dist c-1 to y_1
4	16	5	25
2	4	3	9



e) For an arbitrary c with $c > y_n$, how does $R_{\text{sq}}(c)$ compare to $R_{\text{sq}}(c+1)$? Can you determine which is bigger, and by how much?



Problem 3. Matrix, Vector, Scalar, or Nonsense?

Suppose M is an $m \times n$ matrix, v is a vector in $\mathbb{R}^n$, and s is a scalar. Determine whether each of the following quantities is a matrix, vector, scalar, or nonsense (undefined).

a) Mv

b) vM

c) v^2

d) $M^T M$

e) MM^T

f) $v^T Mv$

g) $(sMv) \cdot (sMv)$

h) $(sv^T M^T)^T$

i) $v^T M^T Mv$

j) $vv^T + M^T M$

Problem 4. Orthogonality

a) Is it possible for a vector to be orthogonal to itself?

dot prod $= 0$ (orthog)

$\vec{v} \cdot \vec{v} = 0$

Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$

$v_1^2 + v_2^2 + v_3^2 + \dots + v_n^2 = 0$

$\Rightarrow \vec{v} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

yes, but only when it's the zero vector

b) Show that if $\vec{u}$ is orthogonal to both $\vec{v}$ and $\vec{w}$, then $\vec{u}$ is also orthogonal to any linear combination of $\vec{v}$ and $\vec{w}$, $\alpha\vec{v} + \beta\vec{w}$.

assume: $\vec{u}$ orthog $\vec{v}$ $(\vec{u} \cdot \vec{v} = 0)$

$\vec{u}$ orthog w $(\vec{u} \cdot \vec{w} = 0)$

need to show:

$\vec{u}$ orthog to $\alpha\vec{v} + \beta\vec{w}$

$(\vec{u} \cdot (\alpha\vec{v} + \beta\vec{w}) = 0)$

proof:

$$\vec{u} \cdot (\alpha\vec{v} + \beta\vec{w})$$

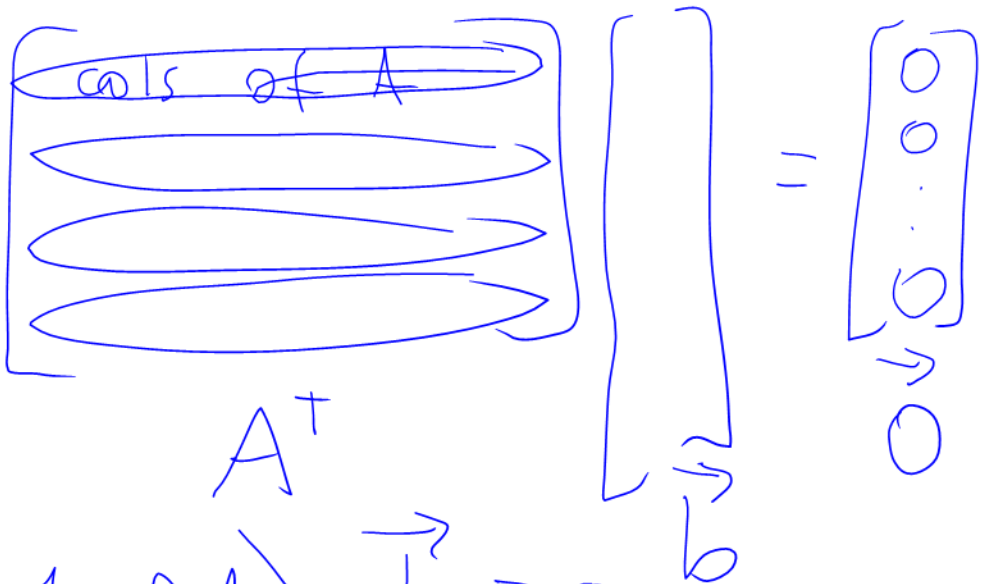
$$= \vec{u} \cdot \alpha\vec{v} + \vec{u} \cdot \beta\vec{w}$$

$$= \alpha\,\vec{u} \cdot \vec{v} + \beta\,\vec{u} \cdot \vec{w}$$

$$= \alpha * 0 + \beta * 0$$

$$= 0$$

c) Show that if $A^T\vec{b} = 0$, then $\vec{b}$ is orthogonal to the **column space** of A, which is the space of all linear combinations of the columns of A.



$(\text{col } 1 \text{ of } A) \cdot \vec{b} = 0$

$(\text{col } 2 \text{ of } A) \cdot \vec{b} = 0$

$\vdots$ same for all cols

$$(SD(x))^2 = (2\,SD(y))^2$$
$$Var(x) = 4\,Var(y)$$

Problem 5. Regression

Suppose you have a dataset

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

where the standard deviation of the x-values, $SD(x)$, is twice the standard deviation of the y-values, $SD(y)$. Let

$$\boxed{y = a + bx}$$

be the regression line with x as the predictor variable and y as the response variable. Let

$$\boxed{x = c + dy}$$

be the regression line with y as the predictor variable and x as the response variable. **Express b in terms of d.**

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$d = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

var(x's)

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

avg of sq. dist to mean

$$\frac{b}{d} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\cancel{n}\cdot Var(y)}{\cancel{n}\cdot Var(x)}$$

$$\frac{b}{d} = \frac{Var(y)}{Var(x)} = \frac{Var(y)}{4\,Var(y)} \Rightarrow \frac{b}{d} = \frac{1}{4}$$

$$\boxed{b = d/4}$$

Same problem, another sol'n

$$SD(x) = 2\,SD(y) \Rightarrow \frac{SD(x)}{SD(y)} = 2$$

$$b = r \cdot \frac{SD(y)}{SD(x)}, \quad d = r \cdot \frac{SD(x)}{SD(y)}$$

$$b = r \cdot \frac{1}{2} \qquad d = r \cdot 2$$

$$\frac{b}{d} = \frac{\cancel{r} \cdot \frac{1}{2}}{\cancel{r} \cdot 2} = \frac{1}{4} \Rightarrow b = d/4$$

Problem 6. Farmfluencer

Billy the avocado farmer heard about the success of 72 year-old Gerald Stratford's viral gardening videos on Twitter and Instagram. After witnessing Gerald turn into the so-called King of Big Veg overnight, Billy is feeling inspired to up his social media game (he's also feeling a little bit jealous).

Billy is new to Instagram and is trying to understand how people gain followers. In particular, he wants to be able to predict the number of followers, y, based on these features:

- number of people they follow, $x^{(1)}$
- number of years since first post, $x^{(2)}$
- average number of posts per day, $x^{(3)}$

a) Suppose Billy has access to a large data set of Instagram accounts, and he uses multiple regression on this data to fit a linear prediction rule of the form

$$H(\vec{x}) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + w_3 x^{(3)}.$$

What does w_2 represent in terms of Instagram followers?

b) What if instead of the number of years since the first post, $x^{(2)}$, Billy instead uses the number of days since the first post, $x^{(4)}$. Now he uses multiple regression to fit a prediction rule of the form

$$H'(\vec{x}) = w_0' + w_1' x^{(1)} + w_3' x^{(3)} + w_4' x^{(4)}.$$

How do the parameters of this prediction rule (w_0', w_1', w_3', w_4') compare to the parameters of original prediction rule (w_0, w_1, w_2, w_3)?

Problem 7. Changing the Prediction Rule

Suppose we have a dataset consisting of variables $x^{(1)}, x^{(2)}$, and y. We use multiple regression to fit a prediction rule of the form

$$H(x^{(1)}, x^{(2)}) = w_0 + w_1(x^{(1)} + x^{(2)}) + w_2 x^{(1)} x^{(2)} + w_3(x^{(1)} + 1)(x^{(2)} + 1) \tag{1}$$

and then again use multiple regression to fit a different prediction rule of the form

$$H'(x^{(1)}, x^{(2)}) = w'_0 + w'_1 x^{(1)} + w'_2 x^{(2)} + w'_3 x^{(1)} x^{(2)}. \tag{2}$$

Which form, (1) or (2), will yield a prediction rule with lower mean squared error? Justify your answer.