RHS: $-3\left(\frac{x_1+x_n}{2}\right)+4$

Mock Exam - Midterm 1

$\frac{\min+\max}{2}$

1. (6 points) Define the extreme mean (EM) of a dataset to be the average of its largest and smallest values. Let

$$f(x) = -3x + 4.$$

$f(x)$

max: $f(x_1)$

min: $f(x_n)$

Show that for any dataset $x_1 \leq x_2 \leq \cdots \leq x_n$,

$$\underbrace{EM(f(x_1), f(x_2), \ldots, f(x_n))}_{\frac{f(x_1)+f(x_n)}{2}} = f(\underbrace{EM(x_1, x_2, \ldots, x_n)}_{f\left(\frac{x_1+x_n}{2}\right)}).$$

LHS:

$$\frac{(-3x_1+4)+(-3x_n+4)}{2}$$

2. (10 points) Consider a new loss function,

$$L(h, y) = e^{(h-y)^2}.$$

Given a dataset $y_1, y_2, \ldots, y_n$, let $R(h)$ represent the empirical risk for the dataset using this loss function.

a) (4 points) For the dataset $\{1, 3, 4\}$, calculate $R(2)$. Simplify your answer as much as possible without a calculator.

b) (6 points) For the same dataset $\{1, 3, 4\}$, perform one iteration of gradient descent on $R(h)$, starting at an initial prediction of $h_0 = 2$ with a step size of $\alpha = \frac{1}{2}$. Show your work and simplify your answer.

3. (8 points) Suppose you have a dataset

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_8, y_8)\}$$

with $n = 8$ ordered pairs such that the variance of $\{x_1, x_2, \ldots, x_8\}$ is 50. Let m be the slope of the regression line fit to this data.

$\text{Var}(X) = \frac{1}{n}\sum_i (x_i - \mu)^2$

Suppose now we fit a regression line to the dataset

$$\{(x_1, y_2), (x_2, y_1), \ldots, (x_8, y_8)\}$$

$\text{Cov}(X,Y) = E(X \cdot Y)$

$\mu_x, \mu_y = 0$

where the first two y-values have been swapped. Let m' be the slope of this new regression line.

If $x_1 = 3$, $y_1 = 7$, $x_2 = 8$, and $y_2 = 2$, what is the difference between the new slope and the old slope? That is, what is $m' - m$? The answer you get should be a number with no variables.

" cross variance "

Hint: There are many equivalent formulas for the slope of the regression line. We recommend using the version of the formula without $\bar{y}$.

$\hat{y} = X\hat{w}$

$$m = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\underbrace{\sum_i (x_i - \bar{x})^2}_{\text{" total variance "}}} \quad ①$$

$$= \frac{\sum_i (x_i - \bar{x}) \cdot y_i}{\sum_i (x_i - \bar{x})^2} \quad ② \ ✓$$

$$= \frac{\sum_i (y_i - \bar{y}) \cdot x_i}{\sum_i (x_i - \bar{x})^2} \quad ③$$

4. (9 points) Consider the dataset shown below.

$x_1 y_1 \quad x_2 y_2$

$$m - m' = \frac{1}{8 \cdot 50} \cdot \left[3\times7 + 8\times2 - (6 + 7\times8) \right]$$

$x_1 y_2 \quad x_2 y_1$

$$21 + 16 - 6 - 56 = -35 + 10 = -25$$

$$= -\frac{25}{8 \cdot 50} = -\frac{1}{16}$$

$$m' - m = \frac{1}{16}$$

X =

$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	y
0	6	8	-5
3	4	5	7
5	-1	-3	4
0	2	1	2

= y

a) (5 points) We want to use multiple regression to fit a prediction rule of the form

$$H(x^{(1)}, x^{(2)}, x^{(3)}) = \overset{w_0 \cdot 1}{w_0} + w_1 x^{(1)} x^{(3)} + w_2 (x^{(2)} - x^{(3)})^2.$$

Write down the design matrix X and observation vector $\vec{y}$ for this scenario. No justification needed.

b) (4 points) For the X and $\vec{y}$ that you have written down, let $\vec{w}$ be the optimal parameter vector, which comes from solving the normal equations $X^T X \vec{w} = X^T \vec{y}$. Let $\vec{e} = \vec{y} - X\vec{w}$ be the error vector, and let e_i be the ith component of this error vector. Show that

$$4e_1 + e_2 + 4e_3 + e_4 = 0.$$

$e \perp \text{column}(X)$

$$X_{\text{new}}: \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{ccc} c & x^{(1)}x^{(3)} & (x^{(2)} - x^{(3)})^2 \\ 1 & 0 & 4 \\ 1 & 15 & 1 \\ 1 & -15 & 4 \\ 1 & 0 & 1 \end{array} \quad \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$[X^T e] = 0$$

$X \in n \times d$

$X^T X = d \times n \times n \times d = d \times d$

$$X^T e = X^T y - X^T \hat{w}$$

$$= X^T y \, 1 - \underbrace{X^T X (X^T X)^{-1}}_{I_d \in \mathbb{R}^{d \times d}} \cdot [\underset{d \times n}{X^T} \underset{n \times 1}{y}] = X^T y - X^T y = \vec{0}$$

$$L(\underset{\substack{\nearrow \\ \text{estimation}}}{h}, \underset{\substack{\nwarrow \\ \text{sample}}}{y}) = e^{(h-y)^2} \quad y = \{1, 3, 4\}$$

$$\underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}}_{I_3}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} = \begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$$

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

y_i	1	3	4	$(h=2)$
$2(h-y_i)$	2	-2	-4	$R(2) = \frac{1}{3}(2e+e^4)$
$(h-y_i)^2$	1	1	4	
$e^{(h-y_i)^2}$	e	e	e^4	

$$\hat{h}_{i+1} = \hat{h}_i - \alpha \cdot R'(h_i)$$ grad descent update rule

$$R(h) = \frac{1}{n}\sum_i e^{(h-y_i)^2} \Rightarrow R'(h) = \frac{1}{n}\sum_i e^{(h-y_i)^2} \cdot 2(h-y_i)$$

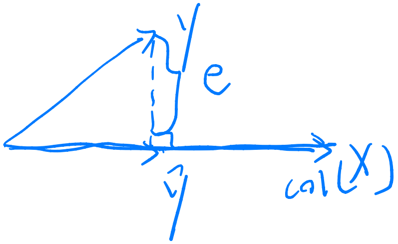
$$\frac{d}{dx}e^x = e^x$$

$$R(2) = \frac{1}{3}(e \cdot 2 + e \cdot -2 + e^4 \cdot -4) = -\frac{4}{3}e^4$$

$$h_{new} = 2 - \frac{1}{2} \cdot (-\frac{4}{3})e^4$$

$$= 2 + \frac{2}{3}e^4$$

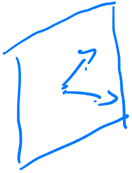
In dis 4:



$e = a - \hat{a}$ or $\hat{a} - a$

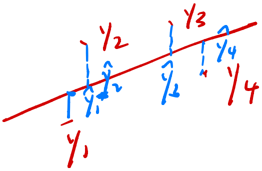
$e \perp b$, $\perp$ space(b)

span(b)



"$y = X w$" $w \in \mathbb{R}^{3\times 1}$

each row is X_i, is one data sample



Over-determined linear system

To "locate" a straight line, I need only 2 point

$\hat{y}$ in column space (X), such that, it acheives

minimized "error"

$e = \hat{y} - y$

Loss: $\|\hat{y}-y\|_2$ ✓

$\sum e^{(\hat{y}-y)^2}$ ✗

$\mathbb{R}^{n\times c}$ ↑ componet / colums.

$y \neq Xw$

$$X^T y = \underbrace{X^T X}_{c\times c} w$$

$$(X^TX)^{-1}X^Ty = \hat{w}$$

$$\hat{y} = X \cdot \hat{w}$$

$$= \underbrace{X(X^TX)^{-1}X^T}_{\text{projection matrix}} y$$

y onto column (X)