Mock Exam - Midterm 1

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

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

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

$$EM(f(x_1), f(x_2), \ldots, f(x_n)) = f(EM(x_1, x_2, \ldots, x_n)).$$

Solution: This linear transformation reverses the order of the data because if $a < b$, then $-3a > -3b$ and so adding four to both sides gives $f(a) > f(b)$. Since $x_1 \leq x_2 \leq \cdots \leq x_n$, this means that the smallest of $f(x_1), f(x_2), \ldots, f(x_n)$ is $f(x_n)$ and the largest is $f(x_1)$. Therefore,

$$\begin{aligned}
EM(f(x_1), f(x_2), \ldots, f(x_n)) &= \frac{f(x_n) + f(x_n)}{2} \\
&= \frac{-3x_n + 4 - 3x_1 + 4}{2} \\
&= \frac{-3x_n - 3x_1}{2} + 4 \\
&= -3\left(\frac{x_1 + x_n}{2}\right) + 4 \\
&= -3EM(x_1, x_2, \ldots, x_n) + 4 \\
&= f(EM(x_1, x_2, \ldots, x_n)).
\end{aligned}$$

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.

Solution: We need to calculate the loss for each data point then average the losses. That is, we need to calculate

$$R(2) = \frac{1}{3}\sum_{i=1}^{3} e^{(2-y_i)^2}.$$

The table below records the necessary information:

y_i	1	3	4
$2 - y_i$	1	-1	-2
$(2 - y_i)^2$	1	1	4
$e^{(2-y_i)^2}$	e	e	e^4

This means

$$\begin{aligned}
R(2) &= \frac{1}{3}\sum_{i=1}^{3} e^{(2-y_i)^2} \\
&= \frac{1}{3}(e + e + e^4) \\
&= \frac{1}{3}(2e + e^4)
\end{aligned}$$

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.

Solution: First, we calculate the derivative of $R(h)$. Using the chain rule, we have

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

To apply the gradient descent update rule, we next have to calculate $R'(h_0)$ or $R'(2)$. Plugging in $h = 2$ to the derivative we calculated above gives

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

The table below records the necessary information (note that we've done most of the work already).

y_i	1	3	4
$2 - y_i$	1	-1	-2
$(2 - y_i)^2$	1	1	4
$e^{(2-y_i)^2}$	e	e	e^4
$e^{(2-y_i)^2} * 2(2 - y_i)$	$2e$	$-2e$	$-4e^4$

Therefore

$$\begin{aligned} R'(2) &= \frac{1}{3}\sum_{i=1}^{3} e^{(2-y_i)^2 * 2(2-y_i)} \\ &= \frac{1}{3}(2e - 2e - 4e^4) \\ &= \frac{-4e^4}{3}. \end{aligned}$$

Applying the gradient descent update rule gives

$$\begin{aligned} h_1 &= h_0 - \alpha * R'(h_0) \\ &= 2 - \frac{1}{2} * \frac{-4e^4}{3} \\ &= 2 + \frac{2e^4}{3} \end{aligned}$$

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.

Suppose now we fit a regression line to the dataset

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

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.

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

Solution: Using the formula for the slope of the regression line, we have

$$\begin{aligned}
m &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{n * Var(x)} \\
&= \frac{(3-\bar{x}) * 7 + (8-\bar{x}) * 2 + \sum_{i=3}^{n}(x_i - \overline{x})y_i}{8 * 50}.
\end{aligned}$$

Note that by interchanging the first two y-values, the terms in the sum from $i = 3$ to n, the number of data points n, and the variance of the x-values are all unchanged. So the slope becomes

$$m' = \frac{(3-\bar{x}) * 2 + (8-\bar{x}) * 7 + \sum_{i=3}^{n}(x_i - \overline{x})y_i}{8 * 50}$$

and the difference between these slopes is given by

$$\begin{aligned}
m' - m &= \frac{(3-\bar{x}) * 2 + (8-\bar{x}) * 7 - ((3-\bar{x}) * 7 + (8-\bar{x}) * 2)}{8 * 50} \\
&= \frac{(3-\bar{x}) * 2 + (8-\bar{x}) * 7 - (3-\bar{x}) * 7 - (8-\bar{x}) * 2}{8 * 50} \\
&= \frac{(3-\bar{x}) * (-5) + (8-\bar{x}) * 5}{8 * 50} \\
&= \frac{-15 + 5\bar{x} + 40 - 5\bar{x}}{8 * 50} \\
&= \frac{25}{8 * 50} \\
&= \frac{1}{16}.
\end{aligned}$$

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

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

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

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

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

Solution: The design matrix X and observation vector $\vec{y}$ are given by

$$X = \begin{bmatrix} 1 & 0 & 4 \\ 1 & 15 & 1 \\ 1 & -15 & 4 \\ 1 & 0 & 1 \end{bmatrix}, \ y = \begin{bmatrix} -5 \\ 7 \\ 4 \\ 2 \end{bmatrix}$$

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

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

Solution: We can rewrite the normal equations in terms of the error vector to get

$$\begin{aligned} X^T X\vec{w} &= X^T\vec{y} \\ \vec{0} &= X^T\vec{y} - X^T X\vec{w} \\ \vec{0} &= X^T(\vec{y} - X\vec{w}) \\ \vec{0} &= X^T\vec{e}. \end{aligned}$$

In particular, since one row of X^T is $\begin{bmatrix} 4 & 1 & 4 & 1 \end{bmatrix}$, when we multiply $\vec{e}$ by this row, the result is zero. This says that $4e_1 + e_2 + 4e_3 + e_4 = 0$, as desired.