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 $x_1 \leq x_2 \leq \cdots \leq x_n$,

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

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.

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 $\overline{y}$.

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.

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

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