Homework 9A: Multiple Linear Regression, Gradients (Theoretical)

EECS 245, Winter 2026 at the University of Michigan

due Friday, March 27th, 2026 at 11:59PM Ann Arbor Time

> Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Homeworks are due to Gradescope by 11:59PM on the due date. See the **syllabus** for details on the slip day policy.
>
> Homework will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should always explain and justify your conclusions, using sound reasoning. Your goal should be to convince the reader of your assertions. If a question does not require explanation, it will be explicitly stated.
>
> Before proceeding, make sure you're familiar with the **collaboration policy**.

Total Points: $10 + 8 + 10 + 8 + 13 + 12 = 61$

Problem 1: Homework 8 Solutions Review (10 pts)

Review the solutions to Homework 8 and pick **two problem parts** (for example, Problem 3c and Problem 5b) from Homework 8 in which your solutions have the most room for improvement, i.e., where they have unsound reasoning, could be significantly more efficient or clearer, etc. **Include a screenshot of your solution to each problem part**, and in a few sentences, explain what was deficient and how it could be fixed.

Alternatively, if you think one of your solutions is significantly better than the posted one, copy it here and explain why you think it is better. If you didn't do Homework 8, choose two problem parts from it that look challenging to you, and in a few sentences, explain the key ideas behind their solutions in your own words.

Problem 2: The Sum of Errors (8 pts)

Consider a set of n points, $(\vec{x}_1, y_1), (\vec{x}_2, y_2), ..., (\vec{x}_n, y_n)$, where each $\vec{x}_i$ is a feature vector in $\mathbb{R}^d$ and each y_i is a scalar.

a) (4 pts) To fit the model

$$h(\vec{x}_i) = w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + ... + w_d x_i^{(d)} = \vec{w} \cdot \text{Aug}(\vec{x}_i)$$

we minimize mean squared error,

$$R(\vec{w}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \vec{w} \cdot \text{Aug}(\vec{x}_i))^2 = \frac{1}{n} \|\vec{y} - X\vec{w}\|^2$$

meaning that $\vec{w}^*$ is chosen to satisfy the normal equations. Explain why the components of the error vector,

$$\vec{e} = \vec{y} - X\vec{w}^*$$

are **guaranteed** to sum to 0.

b) (4 pts) If we decide instead to fit the model

$$h(\vec{x}_i) = w_1 x_i^{(1)} + w_2 x_i^{(2)} + ... + w_d x_i^{(d)} = \vec{w} \cdot \vec{x}_i$$

which has no intercept term, are the components of the error vector $\vec{e} = \vec{y} - X\vec{w}^*$ still guaranteed to sum to 0? If they are, explain why. If they are not, explain why not, but give at least one example dataset where they still do sum to 0.

Problem 3: Moving Things Around (10 pts)

Let X be an $n \times 4$ design matrix whose first column is all 1s, let $\vec{y}$ be an observation vector, and let

$\vec{w}^* = (X^T X)^{-1} X^T \vec{y}$, where $\vec{w}^* = \begin{bmatrix} w_0^* \\ w_1^* \\ w_2^* \\ w_3^* \end{bmatrix}$.

In this problem, you'll reason about modifications to the design matrix and see how they affect the components of $\vec{w}^*$.

a) (3 pts) Let X_a be the design matrix that results from **swapping the first two columns of** X. Let $\vec{v}^* = (X_a^T X_a)^{-1} X_a^T \vec{y}$. Express the components of $\vec{v}^*$ in terms of $w_0^*, w_1^*, w_2^*, w_3^*$.

b) (3 pts) Let X_b be the design matrix that results from **adding 3 to each entry in the *first* column of** X. Let $\vec{v}^* = (X_b^T X_b)^{-1} X_b^T \vec{y}$. Express the components of $\vec{v}^*$ in terms of $w_0^*, w_1^*, w_2^*, w_3^*$.

c) (4 pts) Let X_c be the design matrix that results from **adding 3 to each entry in the *second* column of** X. Let $\vec{v}^* = (X_c^T X_c)^{-1} X_c^T \vec{y}$. Express the components of $\vec{v}^*$ in terms of $w_0^*, w_1^*, w_2^*, w_3^*$.

Problem 4: Gradient Descent Fundamentals (8 pts)

Let $f(\vec{x}) = (x_1 - 5)^2 + (x_1^2 - x_2)^2 + 1$.

a) (4 pts) Find $\nabla f(\vec{x})$, the gradient of $f(\vec{x})$.

b) (4 pts) To minimize $f(\vec{x})$, we'll use gradient descent. Perform one iteration of gradient descent by hand, using the initial guess $\vec{x}^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and learning rate $\alpha = \frac{1}{2}$. What is $\vec{x}^{(1)}$?

Problem 5: Product and Chain Rules (13 pts)

Our goal in this problem is to study the behavior of the function

$$f(\vec{x}) = \frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}}$$

where $x \in \mathbb{R}^n$ and A is a symmetric $n \times n$ matrix (meaning $A = A^T$). This function, called the **Rayleigh quotient**, will play an important role in Chapter 5 of the course, when we eventually study the **dimensionality reduction** problem first introduced in **Chapter 1.1**.

But first, we have to get a handle on a few gradient rules.

a) (4 pts) As described in the **Norm and Chain Rule in Chapter 8.2**, the chain rule for gradients says that if

- $g : \mathbb{R}^d \to \mathbb{R}$ is a **vector**-to-scalar function, and
- $h : \mathbb{R} \to \mathbb{R}$ is a **scalar**-to-scalar function,

then the gradient of the **vector**-to-scalar function $f(\vec{x}) = h(g(\vec{x}))$ is given by

$$\nabla f(\vec{x}) = \left(\frac{\mathrm{d}h}{\mathrm{d}x}(g(\vec{x})) \right) \nabla g(\vec{x})$$

or, perhaps more intuitively,
$$\nabla f(\vec{x}) = h'(g(\vec{x})) \nabla g(\vec{x})$$

Note that we need to pay close attention to the types of functions we're working with. $h(g(\vec{x}))$ is well-defined, but $g(h(\vec{x}))$ is not, since h doesn't take in vectors (it takes in scalars).

Find the gradients of each of the following functions.

  (i) $f_1(\vec{x}) = \log(\vec{x}^T A \vec{x})$, where $\vec{x} \in \mathbb{R}^n$ and A is a symmetric $n \times n$ matrix
  (ii) $f_2(\vec{x}) = e^{-\sin(\vec{a}^T \vec{x})}$, where $\vec{x}, \vec{a} \in \mathbb{R}^n$

Hint: You can use any of the three important gradient rules from Chapter 8.2 without proof.

b) (4 pts) The product rule for gradients is a natural extension of the product rule for derivatives. If $f(\vec{x}) = g(\vec{x})h(\vec{x})$, then

$$\nabla f(\vec{x}) = \nabla(g(\vec{x})h(\vec{x})) = g(\vec{x})\nabla h(\vec{x}) + h(\vec{x})\nabla g(\vec{x})$$

Find the gradients of each of the following functions.

(i) $f_3(\vec{x}) = (\vec{a} \cdot \vec{x})(\vec{b} \cdot \vec{x})$, where $\vec{x}, \vec{a}, \vec{b} \in \mathbb{R}^n$

(ii) $f_4(\vec{x}) = \vec{a}^T \vec{x}\vec{x}^T A\vec{x}$, where $\vec{x}, \vec{a} \in \mathbb{R}^n$ and A is a symmetric $n \times n$ matrix

c) (5 pts) Putting together the chain rule and product rule, show that if

$$f(\vec{x}) = \frac{\vec{x}^T A\vec{x}}{\vec{x}^T \vec{x}}$$

where $\vec{x} \in \mathbb{R}^n$ and A is a symmetric $n \times n$ matrix, then

$$\nabla f(\vec{x}) = \frac{2}{\vec{x}^T \vec{x}}\left(A\vec{x} - f(\vec{x})\vec{x}\right)$$

Problem 6: Convexity (12 pts)

In **this video**, we introduce the formal definition of **convexity** for vector-to-scalar functions. Intuitively, a function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if its graph is a bowl-shaped surface. Formally, f is convex if and only if, for all $\vec{x}$ and $\vec{y}$ in its domain, and for any $t \in [0, 1]$,

$$f(t\vec{x} + (1 - t)\vec{y}) \le tf(\vec{x}) + (1 - t)f(\vec{y})$$

This is a formal way of saying that when you connect any two points on the graph of f with a line segment, the line segment lies on or above the graph of f, never below.

The second derivative test for convexity is more convenient, but it doesn't apply to non-differentiable functions, e.g. $f(x) = |x|$ is convex, but it isn't differentiable.

For each statement below, prove that the statement is true using the formal definition above, or give a counterexample.

a) (4 pts) The sum of two convex functions must also be convex.

b) (4 pts) The difference of two convex functions must also be convex.

c) (4 pts) Suppose $f(x)$ and $g(x)$ are both scalar-to-scalar convex functions and that, for some scalar a, $f(a) = g(a)$. Then, $h(x)$ is also convex, where

$$h(x) = \begin{cases} f(x) & x \le a \\ g(x) & x > a \end{cases}$$

Hint: The statement is false, so focus your energy on finding a counterexample.