# Lab 8: Multiple Linear Regression; The Gradient Vector

EECS 245, Spring 2026 at the University of Michigan

**due** for completion at 11:59PM Ann Arbor Time on Wednesday, June 3rd, 2026

**Name:** _____

**uniqname:** _____

Each lab worksheet will contain several activities, some of which will involve writing code and others that will involve writing math on paper. To receive credit for a lab, you must complete as many of the activities as you can in 2 hours and submit a PDF of your work to Gradescope. We will provide specific instructions on how to submit programming activities (e.g. submitting the notebook or including a screenshot of some output).

**Recap: Multiple Linear Regression (Chapter 7.2)**

Suppose we have $n$ data points, $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots (\vec{x}_n, y_n)$, where each $\vec{x}_i$ is a feature vector of $d$ features:

$$\vec{x}_i = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(d)} \end{bmatrix} \qquad \text{for example, } \vec{x}_i = \begin{bmatrix} \text{height}_i \\ \text{weight}_i \\ \text{shoe size}_i \\ \text{height}_i^2 \\ \cos\left(\text{height}_i \cdot \text{weight}_i\right) \end{bmatrix}$$

This data is stored in the $n \times (d+1)$ matrix $X$, called the **design matrix**, and observation vector $\vec{y}$.

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \ldots & x_1^{(d)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \ldots & x_2^{(d)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \ldots & x_n^{(d)} \end{bmatrix} = \begin{bmatrix} \text{Aug}(\vec{x}_1)^T \\ \text{Aug}(\vec{x}_2)^T \\ \vdots \\ \text{Aug}(\vec{x}_n)^T \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Our goal is to find the optimal parameter vector, $\vec{w}^*$, which minimizes the mean squared error of our model's predictions on the training data.

$$\text{mean squared error} = R_{\text{sq}}(\vec{w}) = \frac{1}{n}||\vec{y} - X\vec{w}||^2$$

The optimal $\vec{w}^*$ satisfies the normal equation, $X^T X \vec{w} = X^T \vec{y}$. To make predictions:

- $\vec{p} = X\vec{w}^*$ is a vector containing the prediction for all $n$ observations.
- $h(\vec{x}_i) = \vec{w}^* \cdot \text{Aug}(\vec{x}_i)$ is the prediction for any one observation $\vec{x}_i$.

## Activity 1: Multiple Linear Regression

Let $X$ be a **full rank** $n \times 3$ design matrix and $\vec{y} \in \mathbb{R}^n$ be an observation vector. Suppose we have already fit a multiple linear regression model of the form

$$h(\vec{x}_i) = w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)}$$

Now, suppose we add the feature $(x_i^{(1)} + x_i^{(2)})$ to our design matrix and train a new model.

**a)** Which of the following are true about the new $n \times 4$ design matrix $X_{\text{new}}$ with our added feature? Select all that apply.

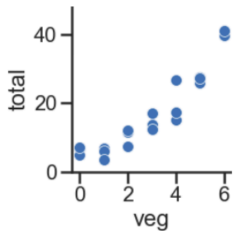
- ☐ The columns of $X_{\text{new}}$ are linearly independent
- ☐ The columns of $X_{\text{new}}$ are linearly dependent
- ☐ $\vec{y}$ is orthogonal to all the columns of $X_{\text{new}}$
- ☐ $\vec{y}$ is orthogonal to all the columns of the original design matrix $X$
- ☐ $\text{colsp}(X) = \text{colsp}(X_{\text{new}})$
- ☐ $X_{\text{new}}^T X_{\text{new}}$ is invertible
- ☐ $X_{\text{new}}^T X_{\text{new}}$ is not invertible

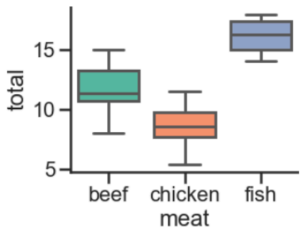
**b)** Find a basis for $\text{nullsp}(X_{\text{new}})$. (This should be quick!)

## Activity 2: Chicken, Beef, or Fish?

Every week, Lauren goes to her local grocery store and buys exactly one pound of meat (either beef, fish, or chicken) but varying amounts of vegetables. We've collected a dataset containing the pounds of vegetables bought, the type of meat bought, and the total bill. Below we display the first few rows of the dataset and two plots generated using the entire (training) dataset.

| veg | meat | total |
|---|---|---|
| 1 | beef | 13 |
| 3 | fish | 19 |
| 2 | beef | 16 |
| 0 | chicken | 9 |




In each part below, we provide you with a model that predicts `total` (her total grocery bill), fit to the dataset by minimizing mean squared error. For each model, determine whether **each optimal parameter $w^*$ is positive, negative or exactly 0**. For example, in part (iv), you'll need to provide 3 answers: one for $w_0^*$, one for $w_1^*$, and one for $w_2^*$.

(i) $h(\vec{x}_i) = w_0$

(ii) $h(\vec{x}_i) = w_0 + w_1 \cdot \text{veg}_i$

(iii) $h(\vec{x}_i) = w_0 + w_1 \cdot \text{meat=chicken}_i$ (one hot encoded feature for chicken)

(iv) $h(\vec{x}_i) = w_0 + w_1 \cdot \text{meat=beef}_i + w_2 \cdot \text{meat=chicken}_i$

(v) $h(\vec{x}_i) = w_0 + w_1 \cdot \text{meat=beef}_i + w_2 \cdot \text{meat=chicken}_i + w_3 \cdot \text{meat=fish}_i$

## Activity 3: Gradients and Partial Derivatives

Suppose $\vec{x} \in \mathbb{R}^3$. Let $g(\vec{x}) = (x_1^2 + x_2 - 3)^2 + (x_1 + x_2^2 - 4)^2 + x_3^2$.

**a)** Find $\nabla g(\vec{x})$. *Hint: Start by finding the partial derivatives of $g$ with respect to $x_1$, $x_2$, and $x_3$.*

**b)** Evaluate $\nabla g \left( \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right)$. The result is a vector in $\mathbb{R}^3$. What does it mean?

**c)** Why is it guaranteed that $g(\vec{x})$ **has** a global minimum?

### Activity 4: The Big Three

In Chapter 8.2, we introduced three key gradient rules for vector-to-scalar functions.

- **Dot product**: If $f(\vec{x}) = \vec{a} \cdot \vec{x}$, then $\nabla f(\vec{x}) = \vec{a}$.
- **Squared norm**: If $f(\vec{x}) = \|\vec{x}\|^2$, then $\nabla f(\vec{x}) = 2\vec{x}$.
- **Quadratic form**: If $f(\vec{x}) = \vec{x}^T A\vec{x}$, then $\nabla f(\vec{x}) = (A + A^T)\vec{x}$.

In each part below, assume $\vec{x}, \vec{a}, \vec{b} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, and $c \in \mathbb{R}$.

**a)** Given $f(\vec{x}) = \vec{x}^T A\vec{x} + \vec{b}^T\vec{x} + c$, find $\nabla f(\vec{x})$.

**b)** Given $f(\vec{x}) = \sum_{i=1}^{n} x_i$, find $\nabla f(\vec{x})$.

**c)** Given $f(\vec{x}) = \|A\vec{x}\|^2$, find $\nabla f(\vec{x})$. *Hint: Use the fact that* $\|\vec{v}\|^2 = \vec{v}^T\vec{v}$.

**d)** Given $f(\vec{x}) = \|\vec{x}\|$, find $\nabla f(\vec{x})$.

**e)** Given $f(\vec{x}) = (\vec{a} \cdot \vec{x})^2$, find $\nabla f(\vec{x})$.
*Hint: Expand $f(\vec{x})$ so that you can use one of the "big three" rules.*

## Activity 5: Quadratic Forms and Symmetry

Suppose $f(\vec{x}) = \vec{x}^T \begin{bmatrix} a & b \\ c & d \end{bmatrix} \vec{x}$, where $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. (If you'd like, as an example, let $A = \begin{bmatrix} 2 & 3 \\ 7 & -8 \end{bmatrix}$.)

(i) Expand $f(\vec{x})$ so that it doesn't involve matrices or vectors.

(ii) Find $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$, and show that $\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$ satisfies the quadratic form gradient rule.

(iii) Discuss: Why do we typically assume that $A$ is symmetric when defining a quadratic form?