Lab 9: Multiple Linear Regression; The Gradient Vector

EECS 245, Winter 2026 at the University of Michigan

due by the end of your lab section

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 all activities and show your lab TA by the end of the lab section.

While you must get checked off by your lab TA **individually**, we encourage you to form groups with 1-2 other students to complete the activities together.

Recap: Multiple Linear Regression (Chapter 7.2)

Suppose we have n data points, $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots (\vec{x}_n, y_n)$, 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 $\vec{y}$.

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

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, $X^TX\vec{w} = X^T\vec{y}$. 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_{new} with our added feature? Select all that apply.

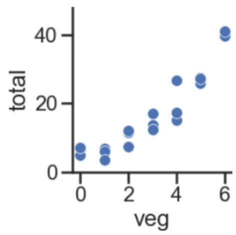
- ☐ The columns of X_{new} are linearly independent
- ☐ The columns of X_{new} are linearly dependent
- ☐ $\vec{y}$ is orthogonal to all the columns of X_{new}
- ☐ $\vec{y}$ is orthogonal to all the columns of the original design matrix X
- ☐ $\text{colsp}(X) = \text{colsp}(X_{\text{new}})$
- ☐ $X_{\text{new}}^T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

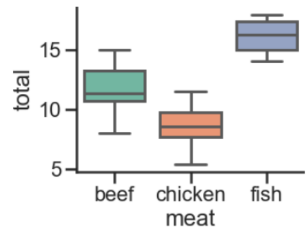
b) Find a basis for $\text{nullsp}(X_{\text{new}})$. (This should be quick!)

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?

The rest of this worksheet is extra practice. Don't feel pressured to answer all of these problems in lab, but make sure to attempt them at some point.

Activity 6: Shuffled Design Matrix

Suppose we want to fit a hypothesis function of the form

$$h(\vec{x}_i) = w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + w_3 (x_i^{(1)})^2 + w_4 (x_i^{(1)} x_i^{(2)})^2$$

Our **full** dataset of $n = 5$ rows is

$x^{(1)}$	$x^{(2)}$	y
1	6	7
-3	8	2
4	1	9
-2	7	5
0	4	6

We know we need to find an optimal parameter vector $\vec{w} \in \mathbb{R}^5$ that satisfies the normal equations. To do so, we build a design matrix, but our columns get all shuffled due to an error with our computer! Our shuffled design matrix is

$$X_{\text{shuffled}} = \begin{bmatrix} 36 & 6 & 1 & 1 & 1 \\ 576 & 8 & -3 & 1 & 9 \\ 16 & 1 & 4 & 1 & 16 \\ 196 & 7 & -2 & 1 & 4 \\ 0 & 4 & 0 & 1 & 0 \end{bmatrix}$$

Let $\vec{s}^* = \begin{bmatrix} s_0^* \\ s_1^* \\ s_2^* \\ s_3^* \\ s_4^* \end{bmatrix}$ be the optimal parameter vector that satisfies the normal equations for this shuffled design matrix. In each part below, select the feature that corresponds to the provided component of $\vec{s}^*$, to "unshuffle" the design matrix.

a) Which term in $h(\vec{x}_i)$ does s_0^* correspond to?

○ intercept ○ $x_i^{(1)}$ ○ $x_i^{(2)}$ ○ $(x_i^{(1)})^2$ ○ $(x_i^{(1)} x_i^{(2)})^2$

b) Which term in $h(\vec{x}_i)$ does s_1^* correspond to?

○ intercept ○ $x_i^{(1)}$ ○ $x_i^{(2)}$ ○ $(x_i^{(1)})^2$ ○ $(x_i^{(1)} x_i^{(2)})^2$

c) Which term in $h(\vec{x}_i)$ does s_2^* correspond to?

○ intercept ○ $x_i^{(1)}$ ○ $x_i^{(2)}$ ○ $(x_i^{(1)})^2$ ○ $(x_i^{(1)} x_i^{(2)})^2$

d) Which term in $h(\vec{x}_i)$ does s_3^* correspond to?

○ intercept ○ $x_i^{(1)}$ ○ $x_i^{(2)}$ ○ $(x_i^{(1)})^2$ ○ $(x_i^{(1)} x_i^{(2)})^2$

Activity 7: Normal Equations Review

Suppose we'd like to fit a linear regression model with an intercept term by minimizing mean squared error, using the full rank design matrix $X \in \mathbb{R}^{n \times (d+1)}$ and observation vector $\vec{y} \in \mathbb{R}^n$.

Let $F = X(X^TX)^{-1}X^T$.

a) How many rows and columns does F have? Give both answers as expressions involving n, d, and/or constants.

In each of the following parts, select one of the following options:

1. The design matrix
2. A vector orthogonal to the column space of X
3. The optimal prediction vector $\vec{p}$
4. The optimal parameter vector $\vec{w}^*$
5. A matrix, vector, or scalar that is none of these options
6. An invalid operation

b) What is XF?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

c) What is FX?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

d) What is $F\vec{y}$?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

e) What is $(I - F)\vec{y}$, where I is the identity matrix with the same dimensions as F?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

f) What is $((I - F)\vec{y}) \cdot \vec{1}_n$, where $\vec{1}_n \in \mathbb{R}^n$ containing all 1's?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6