DSC 40A - Group Work Session 4
due October 28, 2021 at 11:59pm

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. You must work in a group of 2 to 4 students for at least 50 minutes to get credit for this assignment. It's best to join a discussion section if possible.

One person from each group should submit your solutions to Gradescope by 11:59pm on Thursday. Make sure to **tag all group members** so everyone gets credit. This worksheet won't be graded on correctness, but rather on good-faith effort. Even if you don't solve any of the problems, you should include some explanation of what you thought about and discussed, so that you can get credit for spending time on the assignment.

1 Gradient with Respect to a Vector

The derivative of a scalar-valued function $f(\vec{w})$ with respect to a vector input $\vec{w} \in \mathbb{R}^n$ is called the gradient. The gradient of $f(\vec{w})$ with respect to $\vec{w}$, written $\nabla_{\vec{w}} f$ or $\frac{df}{d\vec{w}}$, is defined to be the vector of partial derivatives:

$$\frac{df}{d\vec{w}} = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n}, \end{bmatrix}$$

where $w_1, \ldots, w_n$ are the components of the vector $\vec{w}$. In other words, the gradient of $f(\vec{w})$ with respect to $\vec{w}$ is the same as the gradient of $f(w_1, \ldots, w_n)$, a multivariable function of the components of $\vec{w}$.

Problem 1.

If $\vec{w} \in \mathbb{R}^n$, show that the gradient of $\vec{w}^T\vec{w}$ with respect to $\vec{w}$ is given by

$$\frac{d}{d\vec{w}}(\vec{w}^T\vec{w}) = 2\vec{w}$$

This should remind you of the familiar rule from single-variable calculus that says $\frac{d}{dx}(x^2) = 2x$.

Hint: First, start by writing $\vec{w}^T\vec{w}$ as a sum. What is the partial derivative of that sum with respect to w_1? w_2?

2 Multiple Regression

This problem will check that we're all on the same page when it comes to the notation and basic concepts of regression with multiple features.

Problem 2.

The table below shows the softness and color of several different avocados, which we want to use to predict their ripeness. Each variable is measured on a scale of 1 to 5. For softness, 5 is softest, for color, 5 is darkest, and for ripeness, 5 is ripest.

Avocado	Softness	Color	Ripeness
1	3	4	2.5
2	1	2	2
3	4	5	5

Suppose we have decided on the following prediction rule: given an avocado's softness and color, we average these numbers to produce a predicted ripeness.

a) Is this prediction rule a linear prediction rule or not?

b) Write down the prediction rule as a function $H(\vec{x})$, where

$$\vec{x} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix},$$

with $x^{(1)}$, representing softness and $x^{(2)}$ representing color.

c) Write down the feature vectors $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$ for the first, second, and third avocados in the data set, respectively.

d) Compute the predicted ripeness $H(\vec{x}_1)$, $H(\vec{x}_2)$, $H(\vec{x}_3)$ for each of the three avocados in the data set.

e) Compute the mean squared error of this prediction rule on our data set.

f) Write down the *design matrix*, X.

g) Write down the *parameter vector*, $\vec{w}$ that corresponds to this particular choice of prediction rule. The parameter vector should have three components, one for the bias, and one for each of the features.

h) Check that the entries of $X\vec{w}$ are the predicted ripenesses you found above.

i) Write down the *observation vector* $\vec{y}$.

j) Calculate the length of the vector $X\vec{w} - \vec{y}$.

k) What is the relationship between the length of the vector $X\vec{w} - \vec{y}$ and the mean squared error you found above?