DSC 40A - Group Work Session 4

due Monday, April 29th at 11:59PM

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. **One person** from each group should submit your solutions to Gradescope and **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.

In order to receive full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit. You may not do the groupwork alone.

1 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 1.

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 hypothesis function: given an avocado's softness and color, we average these numbers to produce a predicted ripeness.

a) Is this hypothesis function a linear hypothesis function or not?

b) Write down the hypothesis function 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 hypothesis function 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 hypothesis function. 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?

l) Is the hypothesis function we've used so far, the average of softness and color, the best hypothesis function, or can you find a better one? By better, we mean having a smaller mean squared error. Is there a single best hypothesis function or multiple?

2 Gradient of Vector-Valued Functions

As we start dabbling in multiple linear regression and linear algebra, we will start to work with vector-valued functions. One example of a vector-valued function we have seen in this class is the empirical risk for linear regression.

$$R_{\text{sq}}(\vec{w}) = R_{\text{sq}}\left(\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}\right) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i))^2$$

Recall that to solve for the optimal model parameters, w_0^* and w_1^*, we found the partial derivatives of the empirical risk with respect to each parameter individually. Observe that these make up the components of the gradient of the empirical risk function:

$$\nabla R_{\text{sq}}(\vec{w}) = \begin{bmatrix} \frac{\partial R_{\text{sq}}}{\partial w_0} \\ \frac{\partial R_{\text{sq}}}{\partial w_1} \end{bmatrix}$$

In essence, **the gradient of a vector-valued function is a vector of the function's partial derivatives with respect to each component of the vector**. Notice that this means the dimensions of the function's domain should match the dimensions of the gradient. With regards to the above empirical risk example, R_{sq} takes in a vector of length 2, namely $\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$, so its gradient is also a vector with 2 components. Keep this in mind when solving for gradients.

Problem 2.

Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is the function:

$$f(\vec{x}) = 3x_1^2 + 2x_1x_2 - x_1\cos(x_3)$$

What is the gradient of f? *Hint: Start by finding the partial derivatives of f with respect to each of x_1, x_2, and x_3.*

Problem 3.

Suppose $\vec{a} \in \mathbb{R}^3$, and suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is the function:

$$f(\vec{x}) = \vec{a}^T\vec{x} = a_1x_1 + a_2x_2 + a_3x_3$$

What is the gradient of f? *Hint: Again, start by finding the partial derivatives of f with respect to each of x_1, x_2, and x_3.*

Problem 4.

Perhaps you have begun to notice that some derivative tricks you learned in single variable calculus also apply with vector calculus. We will make some connections to rules you saw in single variable calculus, like the product rule, in this problem.

a) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is the function:

$$f(\vec{x}) = x_1^2 + x_2^2 + ... + x_n^2$$

What is the gradient of f?

b) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is the function:

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

Using the result from Problem 3 and the product rule for single variable functions, what is the gradient of f? *Hint: Recall the product rule for single variable functions,* $\frac{d}{dx} f(x)g(x) = f'(x)g(x) + f(x)g'(x)$.

c) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is the function:

$$f(\vec{x}) = \|\vec{x}\|^2$$

What is the gradient of f?

Note that all three functions f above are really the same function!