DSC 40A - Homework 5

Due: Tuesday, May 16th at 11:59PM

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. You can use a slip day to extend the deadline by 24 hours; you have four slip days to use in total throughout the quarter.

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.

Homeworks should be written up and turned in by each student individually. You may talk to other students in the class about the problems and discuss solution strategies, but you should not share any written communication and you should not check answers with classmates. You can tell someone how to do a homework problem, but you cannot show them how to do it. We encourage you type your solutions in LaTeX, using the Overleaf template on the course website.

For each problem you submit, you should **cite your sources** by including a list of names of other students with whom you discussed the problem. Instructors do not need to be cited.

This homework will be graded out of 52 points. The point value of each problem or sub-problem is indicated by the number of avocados shown.

Note: For full credit, make sure to assign pages to questions when you upload your submission to Gradescope. You will lose points if you don't!

Problem 1. Reflection and Feedback Form

Make sure to fill out this Reflection and Feedback Form, linked here for two points on this homework! This form is primarily for your benefit; research shows that reflecting and summarizing knowledge helps you understand and remember it.

Problem 2. Combinations of Convex Functions

For each statement below, either prove the statement true using the *formal definition* of convexity from Lecture 11, or prove the statement false by finding a concrete counterexample.

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

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

c) If f and g are two convex functions such that $f(a) = g(a)$, where a is some scalar, then the function $h(x)$ is also convex, where:

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

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

Problem 3. Meet the Jensens

As we've seen several times, the variance of a dataset $x_1, x_2, ..., x_n$ is defined:

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

where $\bar{x} = \text{Mean}(x_1, x_2, ..., x_n)$. By expanding the summation (as you did in Homework 4, Problem 2), we find that:

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

Another way of expressing this equation is:

$$\sigma_x^2 = \text{Mean}(x_1^2, x_2^2, ..., x_n^2) - \left(\text{Mean}(x_1, x_2, ..., x_n)\right)^2$$

Since $\sigma_x^2 \geq 0$, this implies that:

$$\begin{aligned} \text{Mean}(x_1^2, x_2^2, ..., x_n^2) - \left(\text{Mean}(x_1, x_2, ..., x_n)\right)^2 &\geq 0 \\ \implies \text{Mean}(x_1^2, x_2^2, ..., x_n^2) &\geq \left(\text{Mean}(x_1, x_2, ..., x_n)\right)^2 \end{aligned}$$

The inequality on the last line can be expressed more generally as:

$$\text{Mean}(g(x_1), g(x_2), ..., g(x_n)) \geq g\left(\text{Mean}(x_1, x_2, ..., x_n)\right)$$

The inequality above is known as Jensen's inequality, and is true for **all convex functions** g. Let's see how we can use Jensen's inequality to prove something useful!

a) Prove that the function $g(x) = -\log x$ is convex. *Hint: Use the second derivative test — it would be difficult to prove this with the formal definition of convexity.*

b) Using Jensen's inequality and $g(x) = -\log x$, prove that, for any dataset of positive numbers $x_1, x_2, ..., x_n$:

$$\frac{x_1 + x_2 + ... + x_n}{n} \geq \left(x_1 \cdot x_2 \cdot ... \cdot x_n\right)^{\frac{1}{n}}$$

The quantity on the left is the familiar arithmetic mean (AM), while the quantity on the right is known as the geometric mean (GM) of $x_1, x_2, ..., x_n$. The entire inequality above is known as the "AM-GM inequality." This is not your first time seeing it — you saw it in Groupwork 2, too!

c) Using Jensen's inequality and a convex function g, prove that the arithmetic mean is greater than or equal to the harmonic mean for any dataset of positive numbers $x_1, x_2, ..., x_n$, i.e.:

$$\frac{x_1 + x_2 + ... + x_n}{n} \geq \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}}$$

Note that you must prove your choice of function g is convex!

Hint: You can use a function that is only convex on an interval, as long as the only inputs you pass into the function come from that interval.

Problem 4. Gradient Descent Gone Wrong

In this problem, we'll familiarize ourselves with how gradient descent works. As mentioned in class, while gradient descent *can* be used to minimize arbitrary differentiable functions, it's most commonly used to find optimal model parameters in the context of empirical risk minimization.

Let's suppose we want to fit a constant model, $H(x) = h$, using degree-3 loss:

$$L_3(y_i, h) = (y_i - h)^3$$

We're given $y_1 = 1, y_2 = 1, y_3 = 7$. Recall, the goal is to find h^*, the best constant prediction for this loss function and dataset.

a) Find $R_3(h)$ and $\frac{dR_3}{dh}(h)$ for this dataset. Both should only involve the variable h; everything else should be constants. (The term "y_i" should not appear in either formula.)

b) Given an initial guess $h_0 = 1$ and a learning rate of $\alpha = \frac{1}{9}$, perform two iterations of gradient descent. What are h_1 and h_2?

c) Is it possible to find an initial guess, h_0, and learning rate, α, for which gradient descent will minimize $R_3(h)$? Why or why not? If we wanted to minimize degree-3 loss, what should we have done instead?

Now that we've gotten a feel for how to use gradient descent to minimize a function of a single variable, we'll use it to minimize a function of multiple variables. Let's suppose we want to fit a simple linear regression model, $H(x) = w_0 + w_1 x$, using **squared loss**. We're searching for the optimal parameters w_0^* and w_1^*, which we can be written in vector form as $\vec{w}^* = \begin{bmatrix} w_0^* \\ w_1^* \end{bmatrix}$.

We're given the following dataset of (x, y) pairs: $\{(1, 5), (2, 7)\}$.

d) Find $R_{\text{sq}}(\vec{w}) = R_{\text{sq}}(w_0, w_1)$ and the gradient vector:

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

Both $R_{\text{sq}}(\vec{w})$ and $\nabla R_{\text{sq}}(\vec{w})$ should only involve the variables w_0 and w_1; everything else should be constants.

e) Given an initial guess $\vec{w}^{(0)} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ and a learning rate of $\alpha = \frac{1}{3}$, perform one iteration of gradient descent. What are the components of $\vec{w}^{(1)}$?

f) In the supplemental Jupyter Notebook, linked **here**, follow the steps mentioned under **Problem 4(f)**. You **do not** need to submit this notebook anywhere. Instead, in your PDF as your answer to this part, include the answers to the prompts in the "Your Job" section.

Problem 5. Trip to Grand Canyon

The second half of this class will be all about probability. To refresh your understanding of probability, we'll work through a problem that only uses concepts from DSC 10.

Utkarsh is planning to go on a road trip to the Grand Canyon with his friends over the Memorial Day weekend. Unfortunately, his friend Varun is worried about the fact that the air there might not be good for his allergies.

The chances that the air is "good" on Friday, Saturday, and Sunday, in two different cities near the Grand Canyon, are given below. The event that the air is good in Phoenix on Saturday depends on the event that the air is good on Friday. All other events are independent.

City	Good on Friday	Good on Saturday	Good on Sunday
Las Vegas	1/3	2/5	3/8
Phoenix	1/4	3/4 if the air was good on Friday 5/9 otherwise	4/5

In all of the parts below, you may leave your answer unsimplified, as long as (1) it's possible to plug your answer directly into a calculator to get the right answer, and (2) you show all of your work. So, an answer like "$1 - \frac{4}{5}$" is acceptable, but "1 minus the probability it is good on Sunday in Phoenix" is not.

a) What is the probability that the air is good in both Las Vegas and Phoenix on Sunday?

b) What is the probability that the air is **not** good in Phoenix on Sunday, given that it was good in Phoenix on both Friday and Saturday?

c) What is the probability that the air is **not** good in Phoenix on Saturday?

d) What is the probability that the air is good on **exactly** two of the three days in Las Vegas?

e) What is the probability that the air was good in Phoenix **at least** once over the three days?