DSC 40A - Homework 5

Due: Friday, Nov 15th 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.

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.

Homework 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 to 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 49 points. The point value and difficulty 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. Mid-quarter Feedback Form

Make sure to fill out this Mid-quarter Reflection and Feedback Form, linked here! Note that this form is anonymous so that you feel comfortable sharing **constructive** feedback. Everyone will be given two points and the course staff is trusting you to fill out the form.

Problem 2. Combinations of Convex Functions

In class, we saw that convex risk functions are nice because they are relatively easy to minimize using gradient descent. But how do we determine if our risk function is convex? One way is to show that it is built from simpler convex functions.

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

a) Suppose that $f_1(x)$ and $f_2(x)$ are convex functions defined on all real numbers. We wish to show that their sum, $f(x) = f_1(x) + f_2(x)$, is also a convex function.

Hint: when proving something like this, first identify the special things that we know about the important entities in the problem. In this case, we know that 1) f is the sum of f_1 and f_2; and 2) f_1 and f_2 are convex functions. We will need to use both pieces of information in our proof. Which should we use first? If you get stuck, ask yourself: have I used all of these pieces of information yet?

Solution: Write your solution here.

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

Solution: Write your solution here.

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.

Solution: Write your solution here.

Problem 3. 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.)

Solution: Write your solution here.

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?

Solution: Write your solution here.

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 did want to minimize degree-3 loss, how must we change the loss function to ensure it works?

Solution: Write your solution here.

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 write 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) Write the empirical risk $R_{\text{sq}}(\vec{w}) = R_{\text{sq}}(w_0, w_1)$ 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.

Solution: Write your solution here.

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 according to the iterative procedure:

$$\vec{w}^{(t)} = \vec{w}^{(t-1)} - \alpha \nabla R(\vec{w}^{(t-1)})$$

What are the components of $\vec{w}^{(1)}$?

Solution: Write your solution here.

f) In the supplemental Jupyter Notebook, linked here, follow the steps mentioned under **Problem 3(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.

Solution: Write your solution here.

Problem 4. 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?

Solution: Write your solution here.

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?

Solution: Write your solution here.

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

Solution: Write your solution here.

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

Solution: Write your solution here.

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

Solution: Write your solution here.

Problem 5. Eventful

a) Let A and B are two events such that $P(B) > 0$. What is the value of $P(A \cap \bar{B}|B)$?

Solution: Write your solution here.

b) Consider two fair 9-sided dice, each with faces numbered 1, 2, 3, ..., 9. Suppose you roll the two dice and look at one of them. You see that this one die is **less than 3**. What is the probability that the sum of the two dice rolls is **greater than 10**?

Solution: Write your solution here.

c) A box contains three cards. One card is red on both sides, one card is green on both sides, and one card is red on one side and green on the other. One card is selected from the box at random, and the color on one side is observed. If this side is green, what is the probability that the other side of the card is also green?

Solution: Write your solution here.

d) In Texas, license plates generally consist of 3 letters followed by 4 numbers. All letters are uppercase, and repeated characters are allowed. ABC-1234 is an example of a Texas license plate.

What is the probability that a randomly generated license plate begins with a vowel or ends in a number divisible by 3? Simplify your answer.

Solution: Write your solution here.

Problem 6. Probability Rules for Three Events

a) The multiplication rule for two events says:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)$$

Use the multiplication rule for two events to prove the multiplication rule for three events:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C|(A \cap B))$$

Hint: You can think of $A \cap B \cap C$ as $(A \cap B) \cap C$.

Solution: Write your solution here.

b) Suppose E, F, and G are events. Explain in words why:

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$$

Intuitively, the relationship between $\cap$ and $\cup$ is similar to the relationship between multiplication and addition; if e, f, g are numbers, then $(e + f) \cdot g = e \cdot g + f \cdot g$ as well.

Solution: Write your solution here.

c) The general addition rule for any two events says:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Use the general addition rule for two events, along with the result of part (b), to prove the general addition rule for three events:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$$

This is often called the "Principle of Inclusion-Exclusion."

Solution: Write your solution here.

d) To identify what students find most important in DSC 10, we want to administer a survey to the students in DSC 20, DSC 30, and DSC 40A. Consider the following information:

- There are 300 students taking at least one of DSC 20, DSC 30, or DSC 40A right now.
- 200 students are taking DSC 20 right now, and 50 students taking DSC 30 right now. There are no students taking both DSC 20 and DSC 30 right now.
- 50 students are taking both DSC 20 and DSC 40A right now, and 30 students are taking both DSC 30 and DSC 40A right now.

Suppose I choose a single student uniformly at random from the population of students taking at least one of DSC 20, DSC 30, and DSC 40A. What is the probability that they are enrolled in DSC 40A? Simplify your answer.

Solution: Write your solution here.