DSC 40A - Group Work Session 1
due Monday, April 8th 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 Minimizers and Maximizers

We've seen that machine learning problems must first be formulated as mathematical problems. Many of these mathematical problems turn out to be optimization problems: finding the value that minimizes or maximizes a function.

For a function of one variable $f(x)$, a value x^* is said to be a **minimizer** of $f(x)$ if

$$f(x^*) \leq f(x) \quad \text{for all } x.$$

Similarly, x^* is said to be a **maximizer** of $f(x)$ if

$$f(x^*) \geq f(x) \quad \text{for all } x.$$

Notice that a function can have multiple minimizers or maximizers. For example, a constant function like $f(x) = 5$ is minimized at all values of x, and it's also maximized at all values of x!

Problem 1.

Should the blank below be filled in with the word *minimizer* or *maximizer* or neither? Prove your result.

If x^* is a minimizer of $f(x)$ then it's a ______________ *of $g(x) = 5f(x) + 3$.*

Problem 2.

Should the blank below be filled in with the word *minimizer* or *maximizer* or neither? Prove your result.

If x^* is a minimizer of $f(x)$ then it's a ______________ *of $g(x) = (f(x))^2$.*

2 Chaining Inequalities

Suppose we have collected a bunch of numbers, $y_1, \ldots, y_n$. Let's assume, too, that these numbers are in sorted order, so that $y_1 \leq y_2 \leq \ldots \leq y_n$.

The *midpoint* of $y_1, \ldots, y_n$ is the average of the smallest and largest number:

$$\text{midpoint} = \frac{y_1 + y_n}{2}.$$

Intuitively, the midpoint is at most y_n and is at least y_1; it lies somewhere in the middle of these two numbers. We can easily prove this with a *chain* of inequalities.

First, we show that the midpoint is at most y_n. We start with the definition:

$$\text{midpoint} = \frac{y_1 + y_n}{2}$$

We can do anything to the right hand side that makes it bigger, keeping in mind that we're trying to get it to look like y_n. Right now there is y_1 hanging out; can we simply change it to a y_n? Yes! Remember that $y_n \geq y_1$, so this would make the right hand side bigger. Therefore, we have to write $\leq$:

$$\leq \frac{y_n + y_n}{2}$$

We can simplify this:

$$= \frac{2y_n}{2}$$

Notice that we wrote $=$ on the last line, not $\leq$. This is because the line is indeed equal to the one before it.

$$= y_n$$

We have made a chain of inequalities and equalities; this one looks like $=, \leq, =, =$. Since $\leq$ is the "weakest link" in the chain, the strongest statement we can make is that the midpoint is $\leq y_n$, but this is what we wanted to say.

Problem 3.

Prove that the midpoint is $\geq y_1$.

Problem 4.

Suppose $y_1, \ldots, y_n$ are all positive numbers. The *geometric mean* of $y_1, \ldots, y_n$ is defined to be:

$$(y_1 \cdot y_2 \cdots y_n)^{1/n}.$$

Prove that the geometric mean is less than or equal to y_n and greater than or equal to y_1 using a chain of inequalities.

3 Empirical Risk Minimization

In class, we've seen how to minimize the empirical risk associated with certain natural loss functions, such as the absolute loss and the squared loss. There are a variety of other possible loss functions we could use instead. This problem explores empirical risk minimization with an alternate choice of loss function.

Problem 5.

In this problem, consider the loss function

$$L(y,h) = \begin{cases} 1, & |y-h| > 1 \\ |y-h|, & |y-h| \le 1 \end{cases}.$$

a) Consider y to be a fixed number. Plot $L(y,h)$ as a function of h.

b) Suppose that we have the following data:

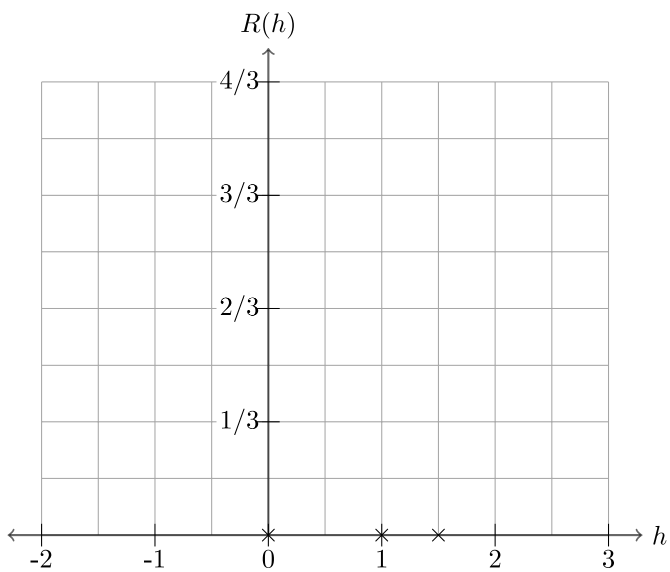
$$\begin{aligned} y_1 &= 0 \\ y_2 &= 1 \\ y_3 &= 1.5 \end{aligned}$$

Plot the empirical risk

$$R(h) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, h)$$

on the domain $[-2, 3]$. It might help to use the grid on the next page; note that the vertical axis tick marks occur in increments of $1/3$ while the horizontal axis tick marks are in increments of 1.

Hint: $R(h)$ is made up of several line segments. What is the slope of each line segment?



Note: You should be able to do this by hand without using technology. After you've done this, click this link to check your work using Desmos, an online graphing calculator.

c) Suppose that we are interested in finding the typical price of an avocado using this loss function. To do so, we have gathered a data set of n avocado prices, $y_1, \ldots, y_n$, and we found the price h^* which minimized the empirical risk (a.k.a, average loss), $R(h) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, h)$.

Unfortunately, a flat tax of c dollars has been imposed on avocados since we performed our analysis, increasing every price in our data set by c.

Is it true that $h^* + c$ is a minimizer of R when we use the new prices, $(y_1 + c), (y_2 + c), \ldots, (y_n + c)$? Explain why or why not by explaining how the graph of R changes.

d) Given avocado prices $\{1/4, 1/2, 3/4, 7/8, 9/8\}$, find a minimizer of R. Provide some justification for your answer.

Hint: You don't need to plot R or do any calculation to find the answer.