DSC 40A - Group Work Session 1

due September 30, 2021 at 11:59pm PT

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 the discussion section on Wednesday if possible.

One person from each group should submit your solutions to Gradescope by 11:59pm PT 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 Summation Notation

You can often verify for yourself if something is true about summation notation by "expanding" the summation symbol and seeing if the property holds. For instance, suppose we want to see if it is true that

$$\sum_{i=1}^{n} c \cdot x_i = c \sum_{i=1}^{n} x_i$$

We start by "expanding" $\sum_{i=1}^{n} c \cdot x_i$:

$$\sum_{i=1}^{n} c \cdot x_i = cx_1 + cx_2 + cx_3 + \ldots + cx_n$$

Now we see that the c can be factored out:

$$\begin{aligned} &= c(x_1 + x_2 + x_3 + \ldots + x_n) \\ &= c \sum_{i=1}^{n} x_i. \end{aligned}$$

This is a simple proof that the property is true. On the other hand, we can prove that a property doesn't hold in the same way: by expanding both sides and showing that they are not equal.

Problem 1.

Show that $\sum_{i=1}^{n}(x_i + y_i) = \left(\sum_{i=1}^{n} x_i\right) + \left(\sum_{i=1}^{n} y_i\right)$.

2 Inequalities

Inequalities are a fundamental part of mathematical proofs. We will go over the basic properties to brush up on things.

- **Law of Trichotomy:** For all $x, y \in \mathbb{R}$, either $x < y$, $x = y$ or $x > y$..
- **Transitive property:** For all $x, y, z \in \mathbb{R}$, if $x \leq y$ and $y \leq z$, then $x \leq z$.

- **Addition property:** For all $x, y, c \in \mathbb{R}$, if $x \leq y$, then $x + c \leq y + c$.
- **Multiplication property:** For all $x, y \in \mathbb{R}$, if $x \leq y$, then
  - $cx \leq cy$ for any $c \geq 0 \in \mathbb{R}$, and
  - $cx \geq cy$ for any $c \leq 0 \in \mathbb{R}$.

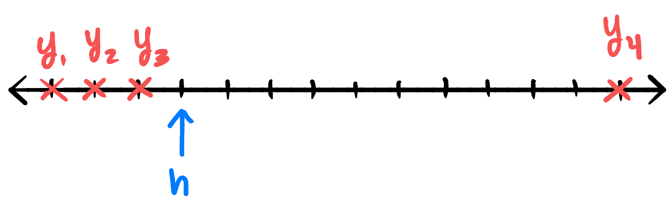
Problem 2.

Suppose $a \leq b$ and $c \leq d$. Which of the statements below are always true? Remember to justify your answers.

1. $a + c \leq b + d$
2. $a - c \leq b + d$
3. $a \leq bc$
4. $ac \leq bd$
5. $|ac| \leq |bd|$
6. $a^2 \leq b^2$
7. $\min(a, c) \leq \min(b, d)$
8. $\min(a, c) \leq \max(b, d)$
9. $\min(a, \max(b, d)) \leq \min(c, \max(b, d))$
10. $\min(a, \max(b, d)) \leq \max(b, d)$

3 Absolute and Squared Error

As we will see in class, the mean's sensitivity to outliers is due to its role as a minimizer of the mean squared error. Now we'll make this more clear. Suppose we have the data $y_1, \ldots, y_n$ drawn below:



For this problem you do not need to know the exact position of the data points, but, if you like, you can assume that the space between each tick mark is one unit and that $y_1 = 1$, $y_2 = 2$, $y_3 = 3$, $y_4 = 14$ and $h = 4$.

Suppose we start out with the prediction h as shown above. There is a tug-of-war going on in the picture above: y_1, y_2, y_3 want h to move closer to them, while y_4 wants h to move to the right to be closer to it. Who wins depends on the choice of error.

Problem 3.

Suppose that absolute error is used (so that we are trying to minimize mean absolute error). Suppose that h is moved one unit to the left. This increases the error for y_4, but decreases the error for y_1, y_2, y_3. Show that the decrease in $|y_1 - h| + |y_2 - h| + y_3 - h|$ makes up for the increase in $|y_4 - h|$ so that moving h to the left decreases the overall error.

Problem 4.

Now suppose that squared error is used. Again suppose that h is moved one unit to the left. Show that the increase in $(y_4 - h)^2$ is larger than the decrease in $(y_1 - h)^2 + (y_2 - h)^2 + (y_3 - h)^2$, so that moving h to the left increases the overall mean squared error.

Informally, moving h to the left always increases the error associated with y_4, whether the absolute error or squared error is used. That is, y_4 always protests against moving h to the left. This protest isn't strong enough in the case of the absolute error, but if the squared error is used, y_4's voice is amplified, and it wins.