DSC 40A - Homework 01

Due: Friday, January 17, 2020

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope on Friday afternoon at 5:00 p.m.

Problem 1.

Which of the following equations involving summation notation are actually wrong? Write the letters of all which are **incorrect**; you do not need to show your work. You can assume that c is a constant, n is a positive integer, k is a positive integer that is less than n, and that $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are real numbers.

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

(b) $$4 \sum_{i=1}^{n} (x_i + y_i) = 4 \sum_{i=1}^{n} x_i + 4 \sum_{i=1}^{n} y_i$$

(c) $$\sum_{i=1}^{n} x_i \cdot y_i = \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)$$

(d) $$\sum_{i=1}^{10} x_i = \sum_{i=1}^{7} x_i + \sum_{i=8}^{10} x_i$$

(e) $$\sum_{i=1}^{n} c = c \cdot n$$

(f) $$\sum_{i=k}^{n} 5 = 5(n-k)$$

(g) $$\sum_{i=1}^{n} x_i = \sum_{j=1}^{n} x_j$$

(h) $$\sum_{i=1}^{n} i = n$$

Solution: The incorrect equations are (c), (f), and (h).

Jaunuary 2020: There was a typo in part (g), so we'll accept your answer if you included (g) as incorrect. It has been fixed above; as it currently appears, it is correct.

A general strategy for determining if the summation notation satisfies a certain property is to assume that n is a small number and to unroll the notation. For instance, if we assume $n = 3$, then:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{3} x_i = x_1 + x_2 + x_3$$

For instance, let's show that (c) is wrong. On the left hand side, we have:

$$\begin{aligned} \sum_{i=1}^{n} x_i \cdot y_i &= \sum_{i=1}^{3} x_i \cdot y_i \\ &= x_1 y_1 + x_2 y_2 + x_3 y_3 \end{aligned}$$

On the other hand, the right hand side is:

$$\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = (x_1 + x_2 + x_3)(y_1 + y_2 + y_3)$$
$$= x_1y_1 + x_1y_2 + x_1y_3 + x_2y_1 + x_2y_2 + x_2y_3 + x_3y_1 + x_3y_2 + x_3y_3$$

Since the left hand side and the right hand side are not the same, this equation is actually wrong.

This actually *proves* that the equation is wrong. If you're trying to prove that something is wrong, its enough to provide a single counterexample, which is what we have done here. To prove that something is right, though, you can't such give a single example.

Equation (f) is wrong; one way to see this is is to take $k = 1$. Then equation (f) is saying that

$$\sum_{i=1}^{n} 5 = 5(n-1) = 5n - 5$$

But we know that

$$\sum_{i=1}^{n} 5 = 5 \cdot n$$

so something is missing. The correct equation is:

$$\sum_{i=k}^{5} = 5(n - k + 1)$$

Equation (h) is wrong. To see this, try $n = 3$. Then:

$$\sum_{i=1}^{3} i = 1 + 2 + 3 = 6.$$

The equation says that this should be 3, so it can't be right.

Problem 2.

In lecture, we argued that a good prediction h is one which has a small mean error:

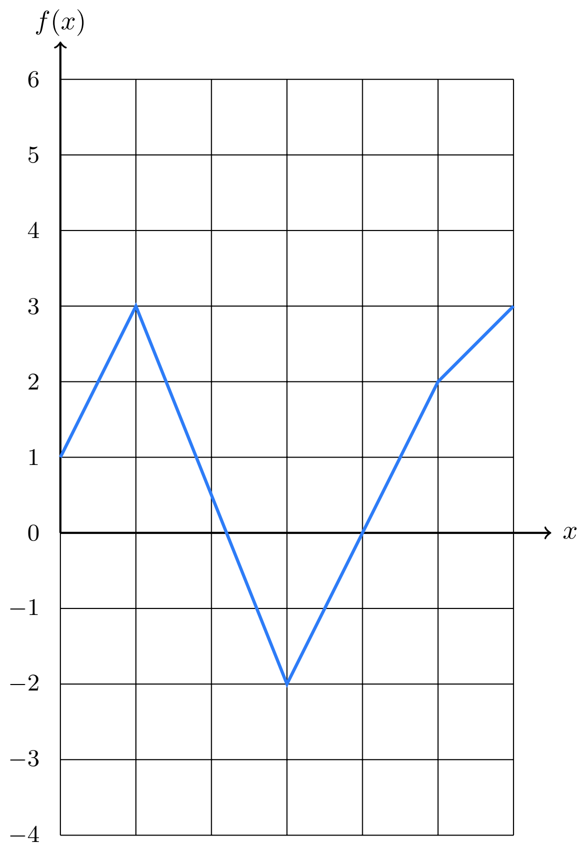
$$R(h) = \frac{1}{n}\sum_{i=1}^{n} |y_i - h|$$

We saw that the median of $y_1, \ldots, y_n$ is the prediction with the smallest mean error. But your friend Zelda thinks that instead of minimizing the mean error, it is better to minimize the *total error*:

$$T(h) = \sum_{i=1}^{n} |y_i - h|$$

In this problem, we'll see if Zelda has a good idea.

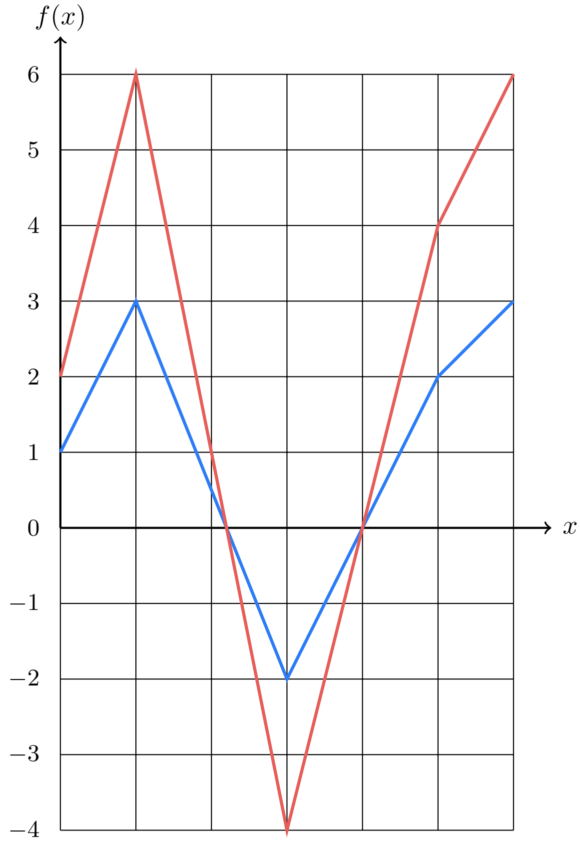
a) Consider the function f plotted below:



Draw the function $g(x) = 2 \cdot f(x)$.

Solution:

The function g is plotted in red below. Notice that the minimizer of g is the same as the minimizer of f.



b) Informally, a *minimizer* of a function f is an input x_{min} where f achieves its minimum value. More formally, x_{min} is a minimizer of f if $f(x_{\text{min}}) \leq f(x)$, no matter what x is.

Suppose that f is some unknown function which takes in a real number and outputs a real number. Suppose that c is an unknown positive constant, and define the function $g(x) = c \cdot f(x)$. Argue that if x_{min} is a minimizer of f, then it is also a minimizer of g.

Hint: what does it mean (formally) for x_{min} to be a minimizer of g? Try to show that $g(x_{\text{min}}) \leq g(x)$, whatever x may be, using a chain of inequalities.

Solution: For x_{min} to be a minimizer of g, then $g(x_{\text{min}}) \leq g(x)$ for every possible x. We'll try to show that this is true.

We'll start with the fact that

$$g(x_{\text{min}}) = c \cdot f(x_{\text{min}})$$

But since x_{min} is a minimizer of f, we have, for every x:

$$\leq c \cdot f(x)$$

We recognize this; it is just $g(x)$:

$$= g(x)$$

We have a chain of inequalities which says that $g(x_{\min}) \leq g(x)$ for every x. This proves that $x_{\min}$ is a minimizer of g.

c) Zelda suggested that we minimize the total error instead of the mean error. What is the minimizer of the total error?

Solution: The minimizer of the total error is also the median. This is because

$$T(h) = n \cdot R(h).$$

Since n is a positive number, the minimizer of $R(h)$ is also a minimizer of $T(h)$. We saw that the median minimizes $R(h)$, so it also minimizes $T(h)$.

As a result, Zelda's idea of minimizing the total error isn't any different than our original idea of minimizing the mean error.

Problem 3.

Suppose that $y_1, \ldots, y_n$ are all real numbers, with y_1 being the smallest and y_n being the largest. Argue that $\text{Mean}(y_1, \ldots, y_n)$ falls somewhere within the interval $[y_1, y_n]$. That is, prove:

$$y_1 \leq \text{Mean}(y_1, \ldots, y_n) \leq y_n.$$

Hint: try to construct two chains of inequalities which show that $\text{Mean}(y_1, \ldots, y_n) \leq y_n$ and $\text{Mean}(y_1, \ldots, y_n) \geq y_1$.

Solution: We'll start by showing that the mean is at most y_n. By definition,

$$\text{Mean}(y_1, \ldots, y_n) = \frac{1}{n} (y_1 + y_2 + \ldots + y_n)$$

Remember that y_n is the largest out of all of these, and that we want to show that this is $\leq y_n$. Here's our trick: if we replace $y_1, y_2, \ldots, y_{n-1}$ by y_n, we get something that is bigger. Hence:

$$=\leq \frac{1}{n} (y_n + y_n + \ldots + y_n)$$

There are n terms in the sum, so:

$$\begin{aligned} &=\leq \frac{1}{n} \cdot n \cdot y_n \\ &= y_n \end{aligned}$$

We have therefore shown that $\text{Mean}(y_1, \ldots, y_n) \leq y_n$. The lower bound of $\text{Mean}(y_1, \ldots, y_n) \geq y_1$ can be proven in the same way, but by replacing each term by y_1 instead of y_n.

Problem 4.

The world's richest person, Jeff Bezos (net worth: \$110 billion), has decided that his true calling is data science and has enrolled in the program here at UCSD. Of course, the data science major is capped at 635 students, and Bezos had to per\$uade someone to drop out so that he could take their place. Assume that he replaced the student who previously had the highest net worth. There were 635 students before Bezos enrolled, and, because one dropped out, there are still 635 after (including Bezos).

a) By how much did the *median* net worth of DSC students increase when Bezos enrolled?

Solution: The median did not change at all. This is easiest to see if we assume that there were only 3 majors, instead of 635. Suppose that before Bezos joined, the net worths of the three students were, in increasing order, w_1, w_2, and w_3. Then w_2 was the median. When Bezos replaced the richest student, the new net worths became (in order):

$$w_1, w_2, \$110 \text{ billion.}$$

So w_2 is still the median.

b) Assume that the student who Bezos replaced had a net worth of \$50,000. By how much did the *mean* net worth of DSC students increase when Bezos enrolled?

Solution:

Let $w_1, \ldots, w_{634}$ be the net worths of the 634 students who were not replaced by Bezos and remember that the net worth of the student who Bezos replaced was \$50,000. Therefore, the old mean net worth was:

$$W_{\text{old}} = \frac{1}{635}\left(w_1 + w_2 + \ldots + w_{634} + \$50{,}000\right)$$

The new net worth is the same, but with 50,000 replaced by Bezos's \$110 billion:

$$W_{\text{new}} = \frac{1}{635}\left(w_1 + w_2 + \ldots + w_{634} + (\$110 \text{ billion})\right)$$

Therefore:

$$\begin{aligned}
W_{\text{new}} - W_{\text{old}} &= \frac{1}{635}\left(w_1 + w_2 + \ldots + w_{634} + (\$110 \text{ billion})\right) - \frac{1}{635}\left(w_1 + w_2 + \ldots + w_{634} + \$50{,}000\right) \\
&= \frac{(\$110 \text{ billion}) - (\$50{,}000)}{635} \\
&= \frac{(\$110 \text{ billion}) - (\$50{,}000)}{635} \\
&= \$173{,}228{,}267.717
\end{aligned}$$

Problem 5.

The National Weather Service of the United States and the Servicio Meteorológico Nacional of Mexico are collaborating to predict the weather at the border between San Diego and Tijuana. To predict the temperature in January, the National Weather Service has collected n temperatures $t_1, \ldots, t_n$ in degrees Fahrenheit and computed the mean and median temperature. But because Mexico is not one of the three countries in the world that still use Fahrenheit, the Servicio Meteorológico Nacional would rather the predicted temperature be stated in degrees Celsius.

For this problem, let $g(t)$ be the function which takes in a temperature in degrees Fahrenheit and outputs the temperature in Celsius. That is, $g(t) = \frac{5}{9} \times (t - 32)$.

a) As chief data scientist at the Servicio Meteorológico Nacional, you're tasked with finding the median temperature in Celsius. Instead of first converting each temperature $t_1, \ldots, t_n$ to Celsius and finding the median of the resulting numbers, you instead simply convert $\text{Median}(t_1, \ldots, t_n)$ to Celsius. Is it true that both approaches give the same result? That is, is it the case that

$$\text{Median}(g(t_1), \ldots, g(t_n)) = g(\text{Median}(t_1, \ldots, t_n))?$$

Give a short justification of your answer. For simplicity, you may assume that there are an odd number of temperatures; this doesn't change the answer.

> **Solution:** It is correct.
>
> The median of a collection of numbers is the middle number (or either of the two middle numbers, if there an even number of numbers). Converting temperatures from Fahrenheit to Celsius doesn't shuffle them; they remain in the same order. So the median of the converted temperatures is simply the converted median.

b) It is indeed true that converting the mean temperature from Fahrenheit to Celsius is the same as converting each temperature to Celsius and finding the mean. That is,

$$\text{Mean}(g(t_1), \ldots, g(t_n)) = g\left(\text{Mean}(t_1, \ldots, t_n)\right)$$

Prove mathematically that this is the case.

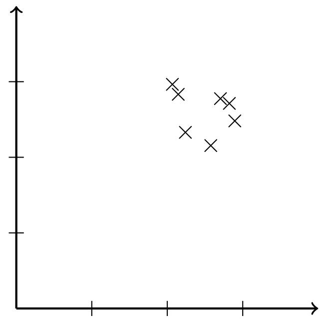
> **Solution:** We'll start with the equation for $\text{Mean}(g(t_1), \ldots, g(t_n))$ and, via some algebra, show that it is the same as $g(\text{Mean}(t_1, \ldots, t_n))$.
>
> We have:
>
> $$\begin{aligned} \text{Mean}(g(t_1), \ldots, g(t_n)) &= \frac{1}{n} \sum_{i=1}^{n} g(t_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{5}{9} \cdot (t_i - 32) \\ &= \frac{5}{9} \cdot \frac{1}{n} \sum_{i=1}^{n} (t_i - 32) \\ &= \frac{5}{9} \left[\frac{1}{n} \sum_{i=1}^{n} (t_i - 32) \right] \\ &= \frac{5}{9} \left[\frac{1}{n} \sum_{i=1}^{n} t_i - \frac{1}{n} \sum_{i=1}^{n} 32 \right] \\ &= \frac{5}{9} \left[\frac{1}{n} \sum_{i=1}^{n} t_i - \frac{1}{n} \cdot 32n \right] \\ &= \frac{5}{9} \left[\frac{1}{n} \sum_{i=1}^{n} t_i - 32 \right] \\ &= \frac{5}{9} \left[\text{Mean}(t_1, \ldots, t_n) - 32 \right] \end{aligned}$$

We recognize this as the formula for converting a temperature in Fahrenheit to Celsius:

$$= g(\text{Mean}(t_1, \ldots, t_n))$$

Problem 6.

SpaceX is trying to land their Falcon 9 rocket on a landing pad, but it keeps missing. Engineers have gathered a list $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ of the coordinates of the previous landings, where $(0, 0)$ is the center of the launchpad. When plotted, the previous landings are distributed as shown below:



The engineers are trying to predict where the next landing will be. Their prediction will be in the form of a pair of numbers, (h_x, h_y), describing the predicted horizontal and vertical position. In order to make a prediction, the engineers choose to minimize the mean squared error of their prediction:

$$\begin{aligned}
\text{MSE}(h_x, h_y) &= \text{mean squared error} \\
&= \frac{1}{n}\sum_{i=1}^{n}(\text{distance between prediction } (h_x, h_y) \text{ and } i\text{th landing } (x_i, y_i))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{(h_x - x_i)^2 + (h_y - y_i)^2}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[(h_x - x_i)^2 + (h_y - y_i)^2\right]
\end{aligned}$$

To minimize $\text{MSE}(h_x, h_y)$, take partial derivatives with respect to h_x and h_y, set both partial derivatives to zero, and solve for h_x and h_y in the resulting system of equations. Show that the prediction which minimizes the mean squared error is:

$$\begin{aligned}
h_x &= \frac{1}{n}\sum_{i=1}^{n} x_i = \text{Mean}(x_1, \ldots, x_n) \\
h_y &= \frac{1}{n}\sum_{i=1}^{n} y_i = \text{Mean}(y_1, \ldots, y_n)
\end{aligned}$$

Solution: We start by taking a partial derivative with respect to h_x. We have:

$$\begin{aligned}
\frac{\partial}{\partial h_x} \mathrm{MSE}(h_x, h_y) &= \frac{\partial}{\partial h_x} \left[\frac{1}{n} \sum_{i=1}^{n} \left((h_x - x_i)^2 + (h_y - y_i)^2 \right) \right] \\
&= \frac{\partial}{\partial h_x} \left[\frac{1}{n} \sum_{i=1}^{n} \left[(h_x - x_i)^2 + (h_y - y_i)^2 \right] \right] \\
&= \frac{1}{n} \cdot \frac{\partial}{\partial h_x} \sum_{i=1}^{n} \left[(h_x - x_i)^2 + (h_y - y_i)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial h_x} \left[(h_x - x_i)^2 + (h_y - y_i)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial}{\partial h_x} (h_x - x_i)^2 + \frac{\partial}{\partial h_x} (h_y - y_i)^2 \right]
\end{aligned}$$

The second partial derivative is zero, since the term contains no h_x. Hence:

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial h_x} (h_x - x_i)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} 2(h_x - x_i)
\end{aligned}$$

Setting this equal to zero and solving:

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} 2(h_x - x_i) = 0 \\
&\implies \frac{2}{n} \sum_{i=1}^{n} (h_x - x_i) = 0 \\
&\implies \frac{2}{n} \sum_{i=1}^{n} h_x - \frac{2}{n} \sum_{i=1}^{n} x_i = 0 \\
&\implies \frac{1}{n} \sum_{i=1}^{n} h_x = \frac{1}{n} \sum_{i=1}^{n} x_i \\
&\implies h_x = \frac{1}{n} \sum_{i=1}^{n} x_i
\end{aligned}$$

The partial derivative with respect to h_y will be nearly the same, except it will be the first term that disappears. Hence $h_y = \frac{1}{n} \sum_{i=1}^{n} y_i$.

Problem 7.

Consider the piecewise function:

$$f(x) = \begin{cases} \frac{1}{2}x^2 + \frac{1}{2}, & |x| \leq 1 \\ |x|, & |x| > 1 \end{cases}$$

We will see next week that this function has a name, and that it plays a role in statistics.

Recall from calculus that a univariate function g is said to be *differentiable* if there is a function g' which gives the slope of f at *every* point. Even though the function f above is piecewise, it is still differentiable, as you will now show.

a) What is the slope of f at any point $x < -1$? Your answer should be a constant.

> **Solution:** The slope is -1. At this point, the piecewise function looks like $-x$.

b) What is the slope of f at any point $x > 1$? Your answer should be a constant.

> **Solution:** The slope is 1. At this point, the function looks like x.

c) Give a formula for the slope of f that works for all x between -1 and 1.

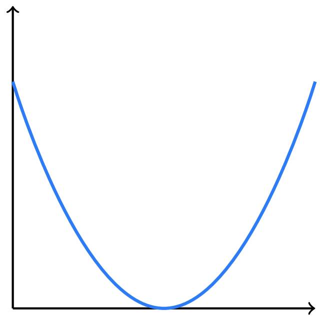
> **Solution:** The slope is x, since the function looks like $\frac{1}{2}x^2$ here.

d) What is $f'(x)$? Your answer should be a piecewise function.

> **Solution:** The slope at any point is given by:
>
> $$f'(x) = \begin{cases} -1, & x < 1 \\ x, & -1 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

Problem 8.

Suppose that f is an unknown function that is shaped like a bowl. For instance, f might be:



Suppose that at a point x_0, the slope of f is negative. Is the minimizer of f to the left of x_0 or to the right of x_0? Why?

Solution: To the right. If the slope of f at x_0 is negative, then the function decreases as we move to the right from x_0. Hence (if the function is bowl-shaped) the minimizer has to be to the right of x_0.