DSC 40A - Homework 2

Due: Friday, January 21, 2022 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.

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.

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 50 points. The point value of each problem or sub-problem is indicated by the number of avocados shown.

Note: For Problem 5, part (c), you'll need to use a supplementary notebook, linked here. We won't grade your code this time; you just need to submit some screenshots of your code's output and a plot, which you can submit along with the rest of the assignment.

Problem 1. Linear Transformations

Suppose we are given a data set $\{d_1, d_2, \ldots, d_n\}$ and know its mean, variance, and standard deviation to be $mean_d$, var_d, and std_d. Consider another data set $\{t_1, t_2, \ldots, t_n\}$, where t_i is a linear transformation of d_i:

$$t_i = f(d_i) = a \cdot d_i + b$$

for each $i = 1, 2, \ldots, n$. Here, a and b are arbitrary constants. Let $mean_t$, var_t, and std_t be the mean, variance, and standard deviation of the transformed data.

a) Express $mean_t$ in terms of $mean_d$, a, and b.

b) Express var_t in terms of var_d, a, and b.

c) Express std_t in terms of std_d, a, and b.

Problem 2. Quadratic Mean

Suppose we are given a data set of size n with $0 < y_1 \leq y_2 \leq \cdots \leq y_n$.

Define a new loss function by

$$L_Q(h, y) = (h^2 - y^2)^2$$

and consider the empirical risk

$$R_Q(h) = \frac{1}{n} \sum_{i=1}^{n} L_Q(h, y_i).$$

a) Show that $R(h)$ has critical points at $h = 0$ and when h equals the **quadratic mean** of the data, defined as

$$QM(y_1, y_2, \ldots, y_n) = \sqrt{\frac{y_1^2 + y_2^2 + \cdots + y_n^2}{n}}.$$

b) Recall from single-variable calculus the **second derivative test**, which says that for a function f with critical point at x^*,

- if $f''(x^*) > 0$, then x^* is a local minimum, and
- if $f''(x^*) < 0$, then x^* is a local maximum.

Use the second derivative test to determine whether each critical point you found in part (a) is a maximum or minimum of $R_Q(h)$.

c) Show that the quadratic mean always falls between the smallest and largest data values, which is a property that any reasonable prediction should have. This amounts to proving the inequality

$$y_1 \leq QM(y_1, y_2, \ldots, y_n) \leq y_n.$$

Problem 3. Happy Family

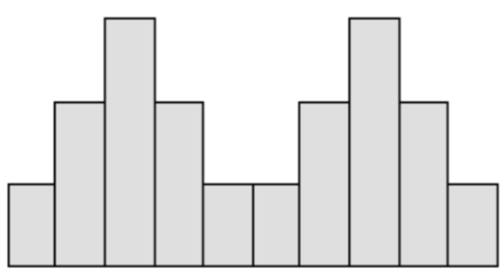
In class, we defined the *mean absolute deviation from the median* as a measure of the spread of a data set. This measure takes the absolute deviations, or differences, of each value in the data set from the median, and computes the mean of these absolute deviations. We can think of this one measure of spread as a member of a family of analogously defined measures of spread:

- mean absolute deviation from the median
- median absolute deviation from the median
- mean absolute deviation from the mean
- median absolute deviation from the mean

While all four of these measures capture the notion of spread, they do so in different ways, and so they may have different values for the same data set.

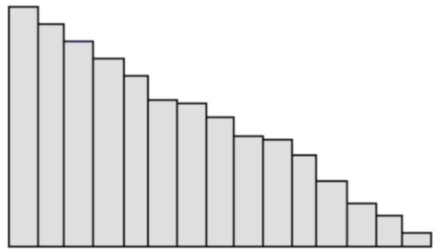
a) For the data set whose histogram is shown below, draw a histogram showing the rough shape of the distribution of the absolute deviations from the mean. Which of these two measures is greater, or are they about the same?

- mean absolute deviation from the mean
- median absolute deviation from the mean



b) For the data set whose histogram is shown below, draw a histogram showing the rough shape of the distribution of the absolute deviations from the median. Which of these two measures is greater, or are they about the same?

- mean absolute deviation from the median
- median absolute deviation from the median



Problem 4. Piecewise Loss

Consider a new loss function,

$$L_p(h, y) = \begin{cases} (h - y)^2, & h \leq y + \frac{1}{2} \\ h - (y + \frac{1}{4}), & h > y + \frac{1}{2} \end{cases}.$$

a) Fix an arbitrary value of y. Draw the graph of $L_p(h, y)$ as a function of h. You should notice that $L_p(h, y)$ is minimized at y.

b) Recall from single-variable calculus the definition of **continuity at a point**:

$$f(x) \text{ is continuous at } x = a \text{ if } \lim_{x \to a} f(x) \text{ exists and is equal to } f(a).$$

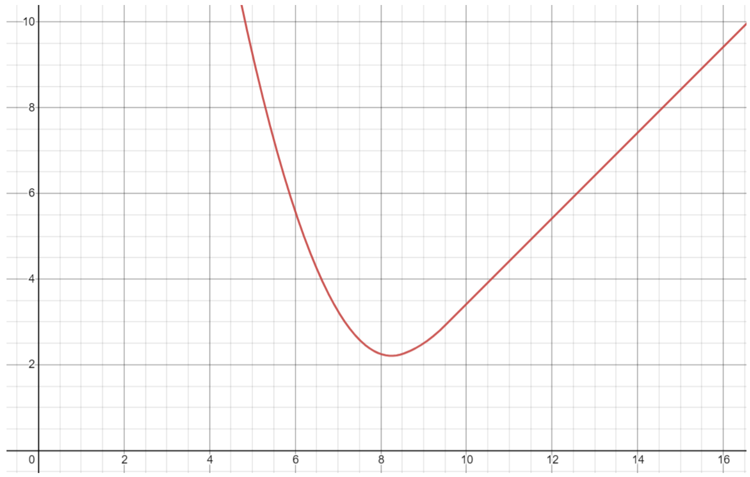
Also, remember that for $\lim_{x \to a} f(x)$ to exist, the left-hand limit and right-hand limit must match:

$$\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x).$$

Fix an arbitrary value of y. Show that as a function of h, $L_p(h, y)$ is continuous for all h.

c) Again fix an arbitrary value of y. The function $L_p(h, y)$ is differentiable for all h. Calculate its derivative $L_p'(h, y)$, which will be a piecewise function of h, and show that both pieces of the function evaluate to the same value at the transition point $h = y + \frac{1}{2}$.

d) Suppose our data set is $\{2, 8, 9\}$. The plot of the empirical risk for this dataset, $R_p(h) = \frac{1}{n} \sum_{i-1}^{n} L_p(h, y_i)$, is shown below:



It is not possible to directly solve for the value of h which minimizes this function. Instead, run gradient descent by hand using an initial prediction of $h_0 = 9$ and a step size of $\alpha = \frac{3}{4}$. Run the algorithm until it converges (it shouldn't take too many iterations). Please show your calculations, and to help the graders track your progress, include a boxed summary with the value of h at each iteration, such as below:

$$\boxed{\begin{array}{l} h_0 = 9 \\ h_1 = \dots \\ h_2 = \dots \\ \vdots \end{array}}$$

e) Does the prediction for the dataset $\{2, 8, 9\}$ seem high or low to you? Why do you think this is the case, based on the graph of the piecewise loss function you drew in part (a)?

Problem 5. Gradient Descent, Linear Regression, and Recipe Nutrition

Soon in the course we are going to learn about linear regression, which is commonly known as finding the "line of best fit". Interestingly enough, the metric used to define "best fit" is a function you are probably quite familiar with already, the squared risk function, also called the mean squared error:

$$R_{sq}(h) = \frac{1}{n}\sum_{i=1}^{n}(y_i - h)^2.$$

So far in this class, our prediction h has been a single real number. When performing linear regression, however, our prediction is allowed to vary with each x_i, according to some linear function $f(x_i) = mx_i + b$, where m and b are real numbers. That means, when talking about how well a linear function "fits" the data, we are measuring the fit by the mean squared error:

$$R_{sq}(f) = R_{sq}(m, b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (mx_i + b))^2$$

and linear regression is just finding the values of m and b that minimize this mean squared error.

While this may look intimidating, we actually have already learned an excellent tool that will be able to help us solve this problem: gradient descent.

Gradient descent is guaranteed to find the global minimum of a function if the function is *convex* (also called concave up) and *differentiable*. In order to verify that gradient descent will indeed find the global minimum for us, we need to prove that the mean squared error is both convex and differentiable. Remember from calculus that a function $g : \mathbb{R} \to \mathbb{R}$ is convex (concave up) if

$$g''(x) = \frac{\mathrm{d}^2 g}{\mathrm{d}x^2} \geq 0, \text{for all } x.$$

a) Use the definition of convexity above to show that

$$R_{sq}(h) = \frac{1}{n}\sum_{i=1}^{n}(y_i - h)^2$$

is a convex function of h.

b) In order to run gradient descent on $R_{sq}(m, b)$, we first need to find its gradient. Thinking of $R_{sq}(m, b)$ as a function of two variables, compute the partial derivatives $\frac{\partial}{\partial m} R_{sq}(m, b)$ and $\frac{\partial}{\partial b} R_{sq}(m, b)$.

c) Now, let's try using all this math to find a regression line for some real data. We'll be looking at nutritional information from recipes on Epicurious and trying to predict the amount of fat based on the number of calories.

In this supplementary notebook (linked), fill in the missing functions based on your answer to part (b), read through the code we've provided, and follow the instructions to submit screenshots of your work. We won't grade your code this time; you just need to submit certain screenshots, which you can submit along with the rest of the assignment.