DSC 40A - Group Work Session 4

due January 24, 2022 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 credit, you must work in a group of two to four students for at least 50 minutes, at one of the scheduled groupwork sessions. You may not do the groupwork alone or meet outside of the scheduled sessions.

1 Least Squares Regression

Recall that the least squares solutions to the problem of fitting a straight line, $h(x) = w_1 x + w_0$, to the data (x_i, y_i) are:

$$w_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$w_0 = \bar{y} - w_1 \bar{x}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Problem 1.

Show that if all the y-values in a data set are multiplied by the same constant c, then the slope of the least squares regression line is also multiplied by c.

Problem 2.

Consider the data set given below. Fit a line $y = a + bx$ by the method of least squares, where x is the predictor variable and y is the response variable. Then, fit a line $x = c + dy$ by the method of least squares, where y is the predictor variable and x is the response variable. You can use a calculator (including Python) to help you find the values of a, b, c, and d, but don't use any tools that perform regression for you.

Are the lines the same or different? Explain why, using the formulas for the slope and intercept of the regression line.

x	0.34	1.38	-0.65	0.68	1.40	-0.88	-0.30	-1.18	0.50	-1.75
y	0.27	1.34	-0.53	0.35	1.28	-0.98	-0.72	-0.81	0.64	-1.59

Problem 3.

Consider the problem of fitting a function of the form $h(x) = b_1 \sin x + b_0$ to the data $(x_1, y_1), \ldots, (x_n, y_n)$. What are the least squares solutions for b_1 and b_0?