DSC 40A - Homework 4

Due: Sunday, May 8, 2022 at 11:59pm PDT

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 PDT on Sunday.

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.

This policy also means that you **should not post or answer homework-related questions on Piazza**, which is a written medium. This includes private posts to instructors. Instead, when you need help with a homework question, talk to a classmate or an instructor in their office hours.

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

Problem 1. Two Views of Linear Regression

Your classmate asked the following interesting question during the class.

Consider the least squares regression on a dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. If you perform the least squares regression using the hypothesis function $y = H(x) = ax + b$, you obtain a prediction rule. On the other hand, if you use an alternative hypothesis function $x = G(y) = cy + d$ and perform least squares regression, you obtain a different prediction rule. Are the two prediction rules the same for the same dataset? In other words, if you use inverse function to compute $y = G^{-1}(x)$, is it the same function as $y = H(x)$?

a) 🥑🥑🥑 Begin this problem by first rewriting function $y = H(x)$ as $y - \bar{y} = H'(x - \bar{x})$ and similarly $x = G(y)$ as $x - \bar{x} = G'(y - \bar{y})$. Here $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. What are the functions H' and G'?

b) 🥑🥑🥑🥑 Using what you obtained, is G' the inverse function of H'? Prove your classmate's conjecture or disprove it by giving a counter example.

Problem 2. Generalizing Matrix Inverse

In linear algebra class, we have studied the solution to the following linear equation:

$$X\vec{w} = \vec{v}. \tag{1}$$

Suppose matrix X is a squared matrix of dimension $d \times d$, vectors $\vec{w}$ and $\vec{v}$ are both of dimension d. Then the number of variables in $\vec{w}$ and the number of equations are both of number d. Solution to the above equation for a full rank matrix X is: $\vec{w} = X^{-1}\vec{v}$.

However, we don't always have the same number of variables and equations. In particular, we sometimes have matrix X to be of dimension $n \times d$, where $n < d$. As a consequence, vector $\vec{w}$ is of dimension d whereas vector $\vec{v}$ is of dimension n.

a) What's the problem with the existing approach?

b) Let's consider a different approach to this problem. If we just wish to find one solution, we can simply look for solutions of the form $\vec{w} = X^T\vec{z}$ (where X^T is the transpose of matrix X). Using this approach, what's the solution to equation (1)? Which matrix needs to be full rank in this case?

Going back to the case where X is a full rank square matrix of dimension $d \times d$, does the current approach lead to a different answer than $\vec{w} = X^{-1}\vec{v}$?

c) Using the approach described above, find a solution to the following equation:

$$\begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 0 & 1 & 1 \end{bmatrix} \vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

where $\vec{w} \in \mathbb{R}^2$.

Problem 3. Vector Calculus Involving Vectors

Let $\vec{v}$ be a fixed vector in $\mathbb{R}^n$. If $\vec{w} \in \mathbb{R}^n$, show that the gradient of $\vec{v}^T\vec{w}$ with respect to $\vec{w}$ is given by

$$\frac{d}{d\vec{w}}(\vec{v}^T\vec{w}) = \vec{v}.$$

This should remind you of the familiar rule from single-variable calculus that says $\frac{d}{dx}(cx) = c$.

Problem 4. Vector Calculus Involving Matrices

Let X be a fixed matrix of dimension $m \times n$, and let $\vec{w} \in \mathbb{R}^n$. In this problem, you will show that the gradient of $\vec{w}^T X^T X \vec{w}$ with respect to $\vec{w}$ is given by

$$\frac{d}{d\vec{w}}(\vec{w}^T X^T X \vec{w}) = 2X^T X \vec{w}.$$

Let $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m$ be the column vectors in $\mathbb{R}^n$ that come from transposing the rows of X. For example, if $X = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 3 & 1 \end{bmatrix}$, then $\vec{r}_1 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}$ and $\vec{r}_2 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$.

a) Show that, for arbitrary $X, \vec{w}$, we can write

$$\vec{w}^T X^T X \vec{w} = \sum_{i=1}^{m} (\vec{r}_i^T \vec{w})^2.$$

Now that we have written

$$\vec{w}^T X^T X \vec{w} = \sum_{i=1}^{m} (\vec{r}_i^T \vec{w})^2,$$

we can apply the chain rule, along with the result of problem 1 above, to conclude that

$$\begin{aligned} \frac{d}{d\vec{w}}(\vec{w}^T X^T X \vec{w}) &= \sum_{i=1}^{m} 2(\vec{r}_i^T \vec{w}) \frac{d}{d\vec{w}}(\vec{r}_i^T \vec{w}) \\ &= \sum_{i=1}^{m} 2(\vec{r}_i^T \vec{w})\vec{r}_i. \end{aligned}$$

b) Next show that, for arbitrary $X, \vec{w}$, we can write

$$2X^T X \overrightarrow{w} = \sum_{i=1}^{m} 2(\overrightarrow{r}_i^T \overrightarrow{w}) \overrightarrow{r}_i.$$

Since you've shown that $\frac{d}{d\overrightarrow{w}}(\overrightarrow{w}^T X^T X \overrightarrow{w})$ and $2X^T X \overrightarrow{w}$ both equal to the same expression, you have proven

$$\frac{d}{d\overrightarrow{w}}(\overrightarrow{w}^T X^T X \overrightarrow{w}) = 2X^T X \overrightarrow{w},$$

as desired.

Problem 5. Intuition for Multivariate Linear Regression

You are given a data set containing information on recently sold houses in San Diego, including

- square footage
- number of bedrooms
- number of bathrooms
- year the house was built
- asking price, or how much the house was originally listed for, before negotiations
- sale price, or how much the house actually sold for, after negotiations

The table below shows the first few rows of the data set. Note that since you don't have the full data set, you cannot answer the questions that follow based on calculations; you must answer conceptually.

House	Square Feet	Bedrooms	Bathrooms	Year	Asking Price	Sale Price
1	1247	3	3	2005	500,000	494,000
2	1670	3	2	1927	1,000,000	985,000
3	716	1	1	1993	335,000	333, 850
4	1600	4	2	1962	830,000	815,000
5	2635	4	3	1993	1,250,000	1,250,000
⋮	⋮	⋮	⋮	⋮	⋮	⋮

a) Suppose you standardize all six variables and fit a linear prediction rule to predict the sale price of the house based on all five of the other variables. Which feature would you expect to have the largest magnitude weight? Without standardizing, which feature would you expect to have the largest magnitude weight? Explain why.

b) Suppose you use multiple linear regression on the original (unstandardized) data and the weight associated with Year is α. Suppose you replace Year with a new predictor variable, Age, which is 0 if the house was built in 2020, 1 if the house was built in 2019, 2 if the house was built in 2018, etc. If we do multiple linear regression again using Age instead of Year, what will be the weight associated with Age in terms of α?

c) Suppose you add a new feature called Rooms, which is the total number of bedrooms and bathrooms in the house. Would multiple linear regression with this extra feature enable you to make better predictions?