DSC 40A - Group Work Session 5
due Monday, Feb 12 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 **include 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 full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit.

1 Multiple Regression

This problem will check that we're all on the same page when it comes to the notation and basic concepts of regression with multiple features.

Problem 1.

The table below shows the softness and color of several different avocados, which we want to use to predict their ripeness. Each variable is measured on a scale of 1 to 5. For softness, 5 is softest, for color, 5 is darkest, and for ripeness, 5 is ripest.

Avocado	Softness	Color	Ripeness
1	3	4	2.5
2	1	2	2
3	4	5	5

Suppose we have decided on the following prediction rule: given an avocado's softness and color, we average these numbers to produce a predicted ripeness.

a) Is this prediction rule a linear prediction rule or not?

Solution: Yes, because the average of softness and color can also be expressed as one half times softness plus one half times color, which is a linear combination of the two features.

b) Write down the prediction rule as a function $H(\vec{x})$, where

$$\vec{x} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix},$$

with $x^{(1)}$, representing softness and $x^{(2)}$ representing color.

Solution:

$$H(\vec{x}) = \frac{x^{(1)} + x^{(2)}}{2} = \frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)}$$

c) Write down the feature vectors $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$ for the first, second, and third avocados in the data set, respectively.

Solution:

$$\vec{x}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \vec{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \vec{x}_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

d) Compute the predicted ripeness $H(\vec{x}_1)$, $H(\vec{x}_2)$, $H(\vec{x}_3)$ for each of the three avocados in the data set.

Solution:

$$\begin{aligned} H(\vec{x}_1) &= \frac{1}{2}x_1^{(1)} + \frac{1}{2}x_1^{(2)} \\ &= \frac{1}{2} * 3 + \frac{1}{2} * 4 \\ &= 3.5 \end{aligned}$$

$$\begin{aligned} H(\vec{x}_2) &= \frac{1}{2}x_2^{(1)} + \frac{1}{2}x_2^{(2)} \\ &= \frac{1}{2} * 1 + \frac{1}{2} * 2 \\ &= 1.5 \end{aligned}$$

$$\begin{aligned} H(\vec{x}_3) &= \frac{1}{2}x_3^{(1)} + \frac{1}{2}x_3^{(2)} \\ &= \frac{1}{2} * 4 + \frac{1}{2} * 5 \\ &= 4.5 \end{aligned}$$

e) Compute the mean squared error of this prediction rule on our data set.

Solution: The squared error of each prediction is:

$$\begin{aligned} (H(\vec{x}_1) - y_1)^2 &= (3.5 - 2.5)^2 = 1 \\ (H(\vec{x}_2) - y_2)^2 &= (1.5 - 2)^2 = 0.25 \\ (H(\vec{x}_3) - y_3)^2 &= (4.5 - 5)^2 = 0.25 \end{aligned}$$

That makes the mean squared error:

$$\frac{1}{3}(1 + 0.25 + 0.25) = 0.5.$$

f) Write down the *design matrix*, X.

Solution:

$$X = \begin{bmatrix} 1 & 3 & 4 \\ 1 & 1 & 2 \\ 1 & 4 & 5 \end{bmatrix}$$

g) Write down the *parameter vector*, $\vec{w}$ that corresponds to this particular choice of prediction rule. The parameter vector should have three components, one for the bias, and one for each of the features.

Solution:

$$\vec{w} = \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

h) Check that the entries of $X\vec{w}$ are the predicted ripenesses you found above.

Solution:

$$X\vec{w} = \begin{bmatrix} 1 & 3 & 4 \\ 1 & 1 & 2 \\ 1 & 4 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 1.5 \\ 4.5 \end{bmatrix}$$

i) Write down the *observation vector* $\vec{y}$.

Solution:

$$\vec{y} = \begin{bmatrix} 2.5 \\ 2 \\ 5 \end{bmatrix}$$

j) Calculate the length of the vector $X\vec{w} - \vec{y}$.

Solution:

$$X\vec{w} - \vec{y} = \begin{bmatrix} 3.5 \\ 1.5 \\ 4.5 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \\ 0.5 \end{bmatrix}$$

The length of this vector is

$$\begin{aligned} \sqrt{1^2 + 0.5^2 + 0.5^2} &= \sqrt{1 + 0.25 + 0.25} \\ &= \sqrt{1.5} \end{aligned}$$

k) What is the relationship between the length of the vector $X\vec{w} - \vec{y}$ and the mean squared error you found above?

Solution: The squared length of the vector $X\vec{w} - \vec{y}$ gives the total squared error. So $\frac{1}{3}$ times the squared length of this vector gives the mean squared error.

$$\frac{1}{3}\sqrt{1.5}^2 = 0.5$$

l) Is the prediction rule we've used so far, the average of softness and color, the best prediction rule, or can you find a better one? By better, we mean having lesser mean squared error. Is there a single best prediction rule or multiple?

Solution: We know that the best prediction rule is one that minimizes the mean squared error, or equivalently, satisfies the normal equations

$$X^T X\vec{w} = X^T\vec{y}.$$

Using the design matrix X from part (f) and the observation vector $\vec{y}$ from part (i), we can calculate

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 4 \\ 4 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 4 \\ 1 & 1 & 2 \\ 1 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 8 & 11 \\ 8 & 26 & 34 \\ 11 & 34 & 45 \end{bmatrix}$$

and

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 4 \\ 4 & 2 & 5 \end{bmatrix} \begin{bmatrix} 2.5 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 9.5 \\ 29.5 \\ 39 \end{bmatrix}$$

so that the best choices of $\vec{w}$ satisfy

$$\begin{bmatrix} 3 & 8 & 11 \\ 8 & 26 & 34 \\ 11 & 34 & 45 \end{bmatrix} \vec{w} = \begin{bmatrix} 9.5 \\ 29.5 \\ 39 \end{bmatrix}.$$

The matrix $X^T X$ appearing on the left side above is not invertible because its third column is the sum of the first two columns. This means there is not a unique solution to the linear system above, but infinitely many solutions.

If we use Gaussian elimination to solve (see steps here), we will find that the general solution takes the form

$$\begin{bmatrix} \frac{11}{14} - x_3 \\ \frac{25}{28} - x_3 \\ x_3 \end{bmatrix}$$

where x_3 is a free variable, meaning it can take on any value. For example, if we let $x_3 = 0$, this says that one optimal prediction rule is

$$H(x) = \frac{11}{14} + \frac{25}{28} \cdot x^{(1)} + 0 \cdot x^{(2)}.$$

In other words, one of the best ways to make predictions is to scale the softness by a factor of $\frac{25}{28}$, add a constant factor of $\frac{11}{14}$ and completely ignore the color!