DSC 40A Fall 2025 - Group Work Session 5
due Monday, October 27th 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 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. You may not do the groupwork alone.

1 Multiple Regression

Suppose we wish to build a model that predicts the ripeness (on a scale of 1 to 5) of a mango in terms of three variables that we can easily observe: the softness, color, and size (also on a scale of 1 through 5). To do this, we go to the grocery store and carefully grade n mangoes (for example, one was softness "1", color "4", size "3", and when we cut into it, a ripeness of "2.5").

In this problem we will create a model to predict the ripeness score using multiple linear regression.

Problem 1.

Going back to the very beginning: review your notes and clearly write out the three steps to the modeling recipe.

Problem 2.

In this problem we will carefully set up each ingredient of the modeling recipe.

a) Suppose we wish to use a linear model to make predictions. Explicitly write out the hypothesis function $H(\vec{x})$, where

$$\vec{x} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \end{bmatrix},$$

with $x^{(1)}$, representing softness, $x^{(2)}$ representing color, and $x^{(3)}$ representing size. What are the parameters of the hypothesis function?

Solution:

$$H(\vec{x}) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + w_3 x^{(3)}$$

Parameters: w_0 (intercept), w_1 for softness, w_2 for color, and w_3 for ripeness.

b) Suppose we have $n \geq 1$ mangoes in our observed dataset, with features $(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n)$. Using **mean squared error**, write down the **empirical risk function** R_{sq} between the hypothesis function above and the observed ripeness values $(y_1, y_2, \ldots, y_n)$. (*Do not use the design matrix yet, simply write the empirical risk directly in terms of the data and parameters.*)

Solution:

$$R_{\text{sq}}(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + w_3 x_i^{(3)}))^2.$$

c) Let X be the **design matrix** for the n mangoes we analyzed and let $\vec{y}$ be the vector of ripeness values for these mangoes. Using your answer from the previous part and some linear algebra, show that

$$R_{\text{sq}}(\vec{w}) = \frac{1}{n}\|y - X\vec{w}\|^2.$$

(*You already saw this formula once or twice in lecture, but you should re-produce those calculations here for practice.*)

Solution:

$$y - X\vec{w} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} x_1^{(1)}w_1 + x_1^{(2)}w_2 + x_1^{(3)}w_3 \\ \vdots \\ x_n^{(1)}w_1 + x_n^{(2)}w_2 + x_n^{(3)}w_3 \end{bmatrix} = \begin{bmatrix} y_1 - (x_1^{(1)}w_1 + x_1^{(2)}w_2 + x_1^{(3)}w_3) \\ \vdots \\ y_n - (x_n^{(1)}w_1 + x_n^{(2)}w_2 + x_n^{(3)}w_3) \end{bmatrix}$$

The length of this vector is

$$\sqrt{(y_1 - (x_1^{(1)}w_1 + x_1^{(2)}w_2 + x_1^{(3)}w_3))^2 + \ldots + (y_n - (x_n^{(1)}w_1 + x_n^{(2)}w_2 + x_n^{(3)}w_3))^2}$$

$$= \sqrt{\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + w_3 x_i^{(3)}))^2}.$$

$$R_{\text{sq}}(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + w_3 x_i^{(3)}))^2$$

$$= \frac{1}{n}\left(\sqrt{\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i^{(1)} + w_2 x_i^{(2)} + w_3 x_i^{(3)}))^2}\right)^2 = \frac{1}{n}\|y - X\vec{w}\|^2.$$

d) Review your notes and find a way to express the optimal parameter $\vec{w}^*$ in terms of X and $\vec{y}$. You do not need to re-produce all of the calculations from lecture here, but you are expected to be comfortable with them.

Solution:

$$\vec{w}^* = (X^T X)^{-1} X^T \vec{y}$$

Problem 3.

The next few parts require direct calculations with the dataset below.

Softness	Color	Size	Ripeness
3	4	3	2.5
1	2	2	2
4	5	2.5	5
3	3.5	3.5	4.5

a) Write down the feature vectors $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$, $\vec{x}_4$ for the mangoes in the data set, in addition to the label vector $\vec{y}$.

Solution:

$$\vec{x}_1 = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix} \qquad \vec{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \qquad \vec{x}_3 = \begin{bmatrix} 4 \\ 5 \\ 2.5 \end{bmatrix} \qquad \vec{x}_4 = \begin{bmatrix} 3 \\ 3.5 \\ 3.5 \end{bmatrix}$$

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

Solution:

$$X = \begin{bmatrix} 1 & 3 & 4 & 3 \\ 1 & 1 & 2 & 2 \\ 1 & 4 & 5 & 2.5 \\ 1 & 3 & 3.5 & 3.5 \end{bmatrix}$$

c) Calculate using a calculator like Wolfram Alpha the matrix $X^T X$ and the vector $X^T \vec{y}$.

Solution:

$$X^T X = \begin{bmatrix} 4 & 11 & \frac{29}{2} & 11 \\ 11 & 35 & \frac{89}{2} & \frac{63}{2} \\ \frac{29}{2} & \frac{89}{2} & \frac{229}{4} & \frac{163}{4} \\ 11 & \frac{63}{2} & \frac{163}{4} & \frac{63}{2} \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 14 \\ 43 \\ \frac{219}{4} \\ \frac{159}{4} \end{bmatrix}$$

d) Finally, use a calculator like Wolfram Alpha to find $(X^T X)^{-1}$. Use your answers in the previous part to write the optimal hypothesis function $H^*(\vec{x})$.

Solution:

$$(X^T X)^{-1} = \begin{bmatrix} \frac{455}{8} & \frac{255}{8} & \frac{-119}{4} & \frac{-53}{4} \\ \frac{255}{8} & \frac{163}{8} & \frac{-75}{4} & \frac{-29}{4} \\ \frac{-119}{4} & \frac{-75}{4} & \frac{35}{2} & \frac{13}{2} \\ \frac{-53}{4} & \frac{-29}{4} & \frac{13}{2} & \frac{7}{2} \end{bmatrix}$$

$$\vec{w}^* = \begin{bmatrix} \frac{455}{8} & \frac{255}{8} & \frac{-119}{4} & \frac{-53}{4} \\ \frac{255}{8} & \frac{163}{8} & \frac{-75}{4} & \frac{-29}{4} \\ \frac{-119}{4} & \frac{-75}{4} & \frac{35}{2} & \frac{13}{2} \\ \frac{-53}{4} & \frac{-29}{4} & \frac{13}{2} & \frac{7}{2} \end{bmatrix} \begin{bmatrix} 14 \\ 43 \\ \frac{219}{4} \\ \frac{159}{4} \end{bmatrix} = \begin{bmatrix} \frac{14081}{8} \\ \frac{10695}{2} \\ \frac{27483}{4} \\ \frac{79867}{16} \end{bmatrix}$$

e) Notice that $|w_3|$ is significantly smaller than $|w_1|$. How can you interpret this observation in the context of the data? What might this tell you about predicting the ripeness of mangoes?

Solution: w_i is the coefficient of feature in coordinate i for $i = 1, 2, 3$ ($i = 0$ corresponds to the intercept). That is, w_1 is the coefficient of the softness feature and w_3 is the coefficient of the size feature. The observation that $|w_3| < |w_1|$ suggests that **softness is more important than size** when building a model to predict ripeness.

f) [**OPTIONAL - PYTHON**] In a Jupyter notebook, use your model $H^*(\vec{x})$ to find the predictions for each of the mangoes in the dataset, and compute the mean squared error.

Solution:

```
import numpy as np

X = np.array([ [1, 3, 4, 3], [1, 1, 2, 2], [1, 4, 5, 2.5], [1, 3, 3.5, 3.5] ])

y = np.array([2.5, 2, 5, 4.5])[:, None]

w_star = (np.linalg.inv(X.T @ X) @ X.T) @ y
pred = X @ w_star

print(pred)
print(f'MSE np.linalg.norm(y - pred)**2 / 4:>1.4f')
```

Output: `[[2.5] [2.  ]  [5.  ]  [4.5]]`
`MSE 0.0000`

2 More on gradients

This problem is a continuation of the last groupwork, and is intended to provide more practice for gradient calculations.

Problem 4.

Each expression on the left is a scalar function of a vector $\vec{x} \in \mathbb{R}^n$, and each expression on the right is a vector field in terms of $\vec{x} \in \mathbb{R}^n$. Draw a line between each function on the left and its gradient on the right. Assume $\vec{a} \in \mathbb{R}^n$ is a fixed vector and $A \in \mathbb{R}^{n \times n}$ is a fixed matrix. *Not all gradients on the right will have a match, and some will have more than one.*

	$(A + A^T)\vec{x}$
$\vec{a}^T\vec{x}$	$A\vec{x}$
$\|\vec{a}^T x\|^2$	$2\vec{a}\vec{a}^T\vec{x}$
$\|\vec{x}\|^2$	$2A^T A\vec{x}$
$\|A\vec{x}\|^2$	$2\vec{x}$
$\vec{x}^T A\vec{x}$	$\frac{1}{2}\vec{a}^T\vec{a}\vec{x}$
$\text{Tr}(\vec{x}\vec{x}^T)$	$\vec{x}^T x$
	$\vec{a}$

As a reminder, Tr is the matrix trace, or the sum of the diagonal elements in a matrix.

Solution:

$$\vec{a}^T\vec{x} = \sum_{i=1}^{n} a_i x_i \implies \nabla \vec{a}^T\vec{x} = \langle a_1, \ldots, a_n \rangle = \vec{a} \tag{1}$$

$$\vec{|a^T\vec{x}|}^2 = \left(\sum_{i=1}^{n} a_i x_i\right)^2 \implies \frac{\partial \vec{|a^T\vec{x}|}^2}{\partial x_i} = 2\vec{a}^T\vec{x}a_i \implies \nabla \vec{|a^T\vec{x}|}^2 = 2\vec{a}\vec{a}^T\vec{x} \tag{2}$$

$$\tag{3}$$

Note that the third option $\|\vec{x}\|^2$ and the sixth option $\mathrm{Tr}(\vec{x}\vec{x}^T)$ are actually the same function, which you encountered in the previous groupwork. Both of their gradients are $2\vec{x}$.

$$\begin{aligned}
\|A\vec{x}\|^2 &= (A\vec{x})^T(A\vec{x}) && (4)\\
&= \vec{x}^T(A^TA)\vec{x} && (5)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \vec{x}^{(i)}(A^TA)_{i,j}\vec{x}^{(j)} && (6)\\
&= \vec{x}^{(1)}(A^TA)_{1,1}\vec{x}^{(1)} + \vec{x}^{(1)}(A^TA)_{1,2}\vec{x}^{(2)} + \ldots + \vec{x}^{(1)}(A^TA)_{1,n}\vec{x}^{(n)} && (7)\\
&+ \vec{x}^{(2)}(A^TA)_{2,1}\vec{x}^{(1)} + \vec{x}^{(2)}(A^TA)_{2,2}\vec{x}^{(2)} + \ldots + \vec{x}^{(2)}(A^TA)_{2,n}\vec{x}^{(n)} && (8)\\
&+ \ldots && (9)\\
&+ \vec{x}^{(n)}(A^TA)_{n,1}\vec{x}^{(1)} + \vec{x}^{(n)}(A^TA)_{n,2}\vec{x}^{(2)} + \ldots + \vec{x}^{(n)}(A^TA)_{n,n}\vec{x}^{(n)}. && (10)
\end{aligned}$$

Therefore, if i is fixed, we can scan the long expression above and find all of the terms that involve $\vec{x}^{(i)}$, finding

$$\frac{\partial \|A\vec{x}\|^2}{\partial x^{(i)}} = \sum_{j=1}^{n} (A^TA)_{i,j}x^{(j)} + (A^TA)_{j,i}x^{(j)} \tag{11}$$

$$\tag{12}$$

but A^TA is symmetric ($(A^TA)^T = A^TA$), so this means $(A^TA)_{i,j}x^{(j)} + (A^TA)_{j,i}x^{(j)} = 2(A^TA)_{i,j}x^{(j)}$. Therefore,

$$\frac{\partial \|A\vec{x}\|^2}{\partial x^{(i)}} = 2\sum_{j=1}^{n} (A^TA)_{i,j}x^{(j)} \tag{13}$$

$$\tag{14}$$

which is the same as matrix multiplication. So $\nabla\|Ax\|^2 = 2A^TA\vec{x}$.

For the fifth term $\vec{x}^TA\vec{x}$, note that a very similar argument applies:

$$\begin{aligned}
\vec{x}^TA\vec{x} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \vec{x}^{(i)}A_{ij}\vec{x}^{(j)} && (15)\\
&= \vec{x}^{(1)}A_{1,1}\vec{x}^{(1)} + \vec{x}^{(1)}A_{1,2}\vec{x}^{(2)} + \ldots + \vec{x}^{(1)}A_{1,n}\vec{x}^{(n)} && (16)\\
&+ \vec{x}^{(2)}A_{2,1}\vec{x}^{(1)} + \vec{x}^{(2)}A_{2,2}\vec{x}^{(2)} + \ldots + \vec{x}^{(2)}A_{2,n}\vec{x}^{(n)} && (17)\\
&+ \ldots && (18)\\
&+ \vec{x}^{(n)}A_{n,1}\vec{x}^{(1)} + \vec{x}^{(n)}A_{n,2}\vec{x}^{(2)} + \ldots + \vec{x}^{(n)}A_{n,n}\vec{x}^{(n)}. && (19)
\end{aligned}$$

Therefore, if i is fixed, we can scan the long expression above and find all of the terms that involve $\vec{x}^{(i)}$, finding

$$\frac{\partial \|A\vec{x}\|^2}{\partial x^{(i)}} = \sum_{j=1}^{n} A_{i,j}x^{(j)} + A_{j,i}x^{(j)} \tag{20}$$

$$\tag{21}$$

Note that $A^T_{i,j} = A_{j,i}$ be definition. So

$$\frac{\partial \|A\vec{x}\|^2}{\partial x^{(i)}} = \sum_{j=1}^{n}(A + A^T)_{i,j}x^{(j)} \tag{22}$$

$$= 2\sum_{j=1}^{n}(A + A^T)_{i,j}x^{(j)} \tag{23}$$

$$\tag{24}$$

Therefore $\nabla \vec{x}^T A\vec{x} = (A + A^T)\vec{x}$.