Data 100, Spring 2023

Discussion #1 Solutions

Welcome to Data 100!

Linear Algebra Fundamentals

1. Linear algebra is what powers linear regression, logistic regression, and PCA (concepts we will be studying in this course). Moving forward, you will need to understand how matrix-vector operations work. That is the aim of this problem.

   Alice, Bob, and Candace are shopping for fruit at Berkeley Bowl. Berkeley Bowl, true to its name, only sells fruit bowls. A fruit bowl contains some fruit and the price of a fruit bowl is the total price of all of its individual fruit.

   Berkeley Bowl has apples for \$2, bananas for \$1, and cantaloupes for \$4. (expensive!). The price of each of these can be written in a vector:

$$\vec{v} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$$

   Berkeley Bowl sells the following fruit bowls:

   1. 2 of each fruit
   2. 5 apples and 8 bananas
   3. 2 bananas and 3 cantaloupes
   4. 10 cantaloupes

   (a) Define a matrix B such that $B\vec{v}$ evaluates to a length 4 column vector containing the price of each fruit bowl. The first entry of the result should be the cost of fruit bowl 1, the second entry the cost of fruit bowl 2, etc.

   Solution:

$$B = \begin{bmatrix} 2 & 2 & 2 \\ 5 & 8 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 10 \end{bmatrix}$$

(b) Alice, Bob, and Candace make the following purchases:

- Alice buys 2 fruit bowl 1s and 1 fruit bowl 2.
- Bob buys 1 of each fruit bowl.
- Candace buys 10 fruit bowl 4s (he really like cantaloupes).

Define a matrix A such that the matrix expression $AB\vec{v}$ evaluates to a length 3 column vector containing how much each of them spent. The first entry of the result should be the total amount spent by Alice, the second entry the amount sent by Bob, etc.

Solution:

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 10 \end{bmatrix}$$

(c) Let's suppose Berkeley Bowl changes their fruit prices, but you don't know what they changed their prices to. Alice, Bob, and Candace buy the same quantity of fruit baskets and the number of fruit in each basket is the same, but now they each spent these amounts:

$$\vec{x} = \begin{bmatrix} 80 \\ 80 \\ 100 \end{bmatrix}$$

In terms of A, B, and $\vec{x}$, determine $\vec{v_2}$ (the new prices of each fruit).

Solution:
We know that $\vec{x} = AB\vec{v_2}$ from the previous part. To solve for $\vec{v_2}$ we need to left-multiply both sides of the above equation by $(AB)^{-1}$. Doing so yields $\vec{v_2} = (AB)^{-1}\vec{x}$.

This assumes that the product AB is invertible. We will explore in the next part why it is invertible.

(d) In the previous part, we assumed that AB (the matrix multiplication of A and B) was invertible. Why is AB (as calculated below) invertible? State two conditions for an arbitrary matrix to be invertible.

$$AB = \begin{bmatrix} 9 & 12 & 4 \\ 7 & 12 & 15 \\ 0 & 0 & 100 \end{bmatrix}$$

Solution: It is full rank since all the vectors are linearly independent. In other words, none of the vectors can be reached using a non-triviail linear combination of the others (or equivalently, we cannot form a non-trivial linear combination of vectors to reach $\vec{0}$), so no vectors are redundant.

We can work this out by observing that only the third vector can travel along the third dimension, so it can neither form either of the two other vectors or be formed by them using a linear combination. The first two vectors point in different directions, so they cannot form the other through a linear combination.

Since the matrix is full rank **and** square, we can invert this matrix. These are the two sufficient conditions for invertibility. Note that there are many equivalent conditions, so this isn't the only solution.

2. As a warm up for the homework, we will introduce matrix inverses and matrix rank.

   - The inverse of a square invertible matrix M, M^{-1} is defined as a matrix such that $MM^{-1} = I$ and $M^{-1}M = I$. The matrix I is a special matrix denoted as the identity matrix where the diagonal elements are 1 and the non-diagonal elements are 0.
   - Linear dependence among a set of vectors $\{v_1, v_2, v_3, ..., v_n\}$ is defined as follows. If any (non-trivial) linear combination of the vectors can produce the zero vector, then the set of vectors is linearly dependent.

     In other words, if we can multiply the vectors v_i with some scalar α_i and sum the quantity to obtain the zero vector (given at least one $\alpha_j \neq 0$, then the set is linearly dependent.

     $$\sum_{i=1}^{n} \alpha_i v_i = 0 \text{ such that some } \alpha_j \neq 0 \implies \text{linear dependence}$$

     Any set of vectors such that we cannot obtain the zero vector as described above is linearly independent.
   - The (column) rank of a matrix M is the maximal number of linearly independent column vectors in M. A full rank matrix has a column rank equal to the number of column vectors.

   We will go over all of these definitions applied to relevant practical examples in the following subparts.

   (a) Consider the matrix $M = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ containing two column vectors $v_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$. Is it possible to construct the zero vector using a linear combination of the column vectors? What can be concluded about the rank of the matrix M?

> **Solution:** No, it is impossible to construct the zero vector if we use at least one $\alpha_i \neq 0$ since the first vector can't affect the second dimension and vice versa. Hence, neither of the vectors can "undo" each other. As a more formal proof:
>
> $$\alpha_1 v_1 + \alpha_2 v_2 = \begin{bmatrix} 2\alpha_1 \\ 3\alpha_2 \end{bmatrix}$$
>
> If at least one of the α values are not 0, then the above quantity can never be 0. Hence, this matrix is full rank (i.e. the set of vectors is linearly independent).

(b) Consider the inverse matrix $M^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ of M. Carry out the matrix multiplication MM^{-1}, and determine what M^{-1} must be.

> **Solution:** We carry out the matrix multiply:
>
> $$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 2a & 2b \\ 3c & 3d \end{bmatrix}$$
>
> Hence, since we know that $MM^{-1} = I$, $b = c = 0$ and $2a = 1$ and $3d = 1$. Thus, the inverse matrix is:
>
> $$M^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$
>
> An interesting fact about diagonal matrices (i.e. matrices that are non-zero on the diagonal entries but zero everywhere else) is that their inverse is simply the elementwise multiplicative inverse of the diagonal entries!

(c) Consider a different matrix $Q = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 5 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$. What is the column rank of the matrix? Is the matrix invertible?

> **Solution:** We can construct the zero vector with all of v_1, v_2 and v_3, so this initial set is linearly dependent since $5v_1 + 5v_2 - v_3 = 0$. Hence, the maximal number of linearly independent vectors we can have is 2, where we remove any one of the vectors among v_1, v_2 and v_3. A similar argument as from the first subpart can be applied to explain why there is no linear dependence in any 2 of the vectors (i.e. $\{v_1, v_2\}$, $\{v_2, v_3\}$, and so on). The column rank is 2.
>
> The matrix is not invertible since it is not square.

(d) Consider a matrix R, which is equal to the transpose of the matrix Q: $R = Q^T$. What is the column rank of the matrix R? Is the matrix R invertible?

Solution: We take the transpose:

$$R = Q^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 5 & 5 \end{bmatrix}$$

The column rank is 2 because we can never "undo" the first vector dimension with the second and vice versa, similarly to previous subparts. Thus, this matrix is full column rank.

The matrix is not invertible since it is not square, but take a look at the next bonus question to find out the conditions under which the matrix $R^T R$ can be inverted (this figures into how we study linear regression later on)!

Calculus

In this class, we will have to determine which inputs to a functions minimize the output (for instance, when we choose a model and need to fit it to our data). This process involves taking derivatives.

In cases where we have multiple inputs, the derivative of our function with respect to one of our inputs is called a *partial derivative*. For example, given a function $f(x, y)$, the partial derivative with respect to x (denoted by $\frac{\partial f}{\partial x}$) is the derivative of f with respect to x, taken while treating all other variables as if they're constants.

3. Suppose we have the following scalar-valued function on x and y:

$$f(x, y) = x^2 + 4xy + 2y^3 + e^{-3y} + \ln(2y)$$

(a) Compute the partial derivative of $f(x, y)$ with respect to x.

Solution:

$$\frac{\partial}{\partial x} f(x, y) = 2x + 4y$$

(b) Compute the partial derivative of $f(x, y)$ with respect to y.

Solution:

$$\frac{\partial}{\partial y} f(x, y) = 4x + 6y^2 - 3e^{-3y} + \frac{2}{2y}$$

(c) The gradient of a function $f(x, y)$ is a vector of its partial derivatives. That is,

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}^T$$

$\nabla f(x, y)$ tells us the magnitude and direction in which f is moving, at point (x, y). This is analogous to the single variable case, where $f'(x)$ is the rate of change of f, at the point x.

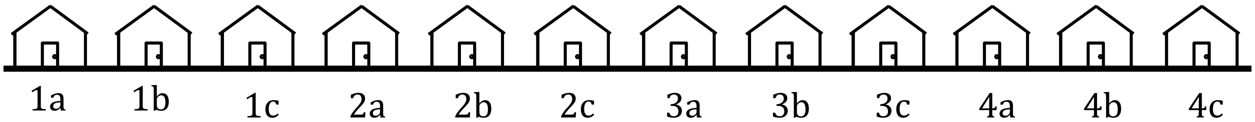
Using your answers to the above two parts, compute $\nabla f(x, y)$ and evaluate the gradient at the point $(x = 2, y = -1)$.

Solution:

$$\nabla f(x, y) = [2x + 4y, 4x + 6y^2 - 3e^{-3y} + \frac{2}{2y}]^T$$

$$\nabla f(2, -1) = [0, 13 - 3e^3]^T$$

Probability & Sampling



4. Kalie wants to measure interest for a party on her street. She assigns numbers and letters to each house on her street as illustrated above. She picks a letter "a", "b", or "c" at random and then surveys every household on the street ending in that letter.

   (a) What is the chance that two houses next door to each other are both in the sample?

   > **Solution:** None of the adjacent houses end in the same letter, so the chance is zero.

   (b) Now, suppose that Kalie decides to collect an SRS of one house instead. What is the probability that house 1a is **not** selected in Kalie's SRS of one house?

   > **Solution:** The probability that house 1a is selected is $\frac{1}{12}$. Therefore, the probability that house 1a is not selected is $1 - \frac{1}{12} = \frac{11}{12}$.

   (c) Kalie decides to collect a SRS of four houses instead of a SRS of one house. What is the probability that house 1a is **not** in Kalie's simple random sample of four houses?

   > **Solution:** This time, we are taking a sample of 4 houses. But, we can apply a similar approach from part b to determine the probability of missing house 1a in each of the four selected houses. Then we multiply the four probabilities together to get our answer. The probability that house 1a is not in Kalie's sample is $\frac{11}{12} \cdot \frac{10}{11} \cdot \frac{9}{10} \cdot \frac{8}{9} = \frac{8}{12} = \frac{2}{3}$.

   (d) Instead of surveying every member of each house from the SRS of four houses, Kalie decides to only survey two members in each house. Four people live in house 1a, one of whom is Bob. What is the probability that Bob is **not** chosen in Kalie's new sample?

   > **Solution:** The probability that house 1a is included in Kalie's initial SRS is $\frac{1}{3}$. Given that house 1a is selected, the probability that Bob is one of the two people

surveyed is $\frac{1}{2}$. Therefore, the probability that Bob *is* surveyed is $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$. Thus the probability that Bob is **not** selected is $1 - \frac{1}{6} = \frac{5}{6}$.