Data 100, Summer 2020

Discussion #1

June 22, 2020

Welcome to Data 100!

Linear Algebra Fundamentals

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

   Joey, Deb, and Sam 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.

(b) Joey, Deb, and Sam make the following purchases:

- Joey buys 2 fruit bowl 1s and 1 fruit bowl 2.
- Deb buys 1 of each fruit bowl.
- Sam 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 Joey, the second entry the amount sent by Deb, etc.

(c) Let's suppose Berkeley Bowl changes their fruit prices, but you don't know what they changed their prices to. Joey, Deb, and Sam 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).

Calculus

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

2. 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.

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

(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 $(x = 2, y = -1)$.

Proportions

In Data 100 you will typically work with multiple variables and large data sets. But before we get carried away by complexity, let's make sure we have our feet on the ground when it comes to interpreting simple quantities like proportions.

3. Investigators at the scene of a crime find a footprint that shows a distinctive pattern on the sole. They identify the type of shoe, and then they find a person owns that kind of shoe and could have committed the crime. They put this person on trial for the crime.

   After looking at sales patterns and so on, the investigators find that of the 10,000 other people who could have committed the crime, 1 in 1,000 own that kind of shoe.

   The prosecution says that given these findings, the chance that the defendant is not the guilty person is 1 in 1000.

   The prosecution has made an error called the "prosecutor's fallacy." Unfortunately it's rather common. Let's see what the error is and what conclusions we can draw from the evidence.

   (a) There are 10,001 people who could have committed the crime. Define a person to be "Consistent with Evidence" if the person owns the kind of shoe identified by the investigators. Fill in the table below with the counts of people in the four categories. The four counts should add up to 10,001, and you should assume, as the prosecution did, that only one person is guilty.

	Guilty	Not Guilty
Consistent with Evidence		
Not Consistent with Evidence		

   (b) The prosecution has reported a proportion as a chance. Whether they know it or not, this implies they are assuming that the defendant is like a person drawn at random from the group who could have committed the crime. So let's assume that too. That is, we assume the defendant is drawn at random from 10,001 people of whom 1 is guilty.

   Use the table in Part **a** to fill in the blanks with choices from among "Guilty", "Not Guilty", "Consistent with Evidence", and "Not Consistent with Evidence". The vertical bar is the usual notation for "given".

   Under this assumption, $\frac{1}{1000} = P(\_\_\_\_\_\_\_ \mid \_\_\_\_\_\_\_)$.

(c) What the investigators know is that the defendant has the fateful type of shoe. Fill in the blanks:

Given the findings of the investigators, the chance that the defendant is not guilty is $P(\_\_\_\_\_\_\_ \mid \_\_\_\_\_\_\_) = \_\_\_\_\_\_\_$.

The last blank should be filled with a fraction, and the first two should be filled choices from among "Guilty", "Not Guilty", "Consistent with Evidence", and "Not Consistent with Evidence".

Note: The prosecution's error is to confuse the probabilities in Parts **b** and **c**.