Written homework 3

Math 187: Introduction to Applied Linear Algebra

Due in class: Wednesday, September 18

1. Problem 1.7 on page 26.
2. Problem 1.11 on page 26.
3. Read section 1.5 to prepare to answer these questions about computational complexity. All of our vector computations have simple parts, but the number of simple operations required to compute a vector operation can be taxing for a computer. In the worst case, it takes too much time for even a computer to compute, and we are out of luck. Let's start keeping track of how computationally complicated our operations can get. Let x and y be vectors in $\mathbb{R}^{20}$. How many multiplications and additions does it take to compute the following?
   (a) $\mathbf{1}^T(x+y)$
   (b) $\mathbf{1}^T x + \mathbf{1}^T y$
   (c) $(x^T y)y$
   (d) $\frac{y^T x}{y^T y}\, y$
4. The file `candy-data.csv`, available on Canvas, contains a collection of attribute data for 85 different types of candy along with a type of score collected from internet voting. Here is a description of the data from the group that collected it, FiveThirtyEight:

   > What's the best (or at least the most popular) Halloween candy? That was the question this dataset was collected to answer. Data were collected by creating a website where participants were shown presenting two fun-sized candies and asked to click on the one they would prefer to receive. In total, more than 269 thousand votes were collected from 8,371 different IP addresses.

   In this problem, you will use your own taste preferences to rank these candies based on their attributes and then compare your ranking with the public's internet voting. You'll get more practice on importing a csv file into R and selecting row vectors and column vectors.

   (a) Describe how you will create candy bar vectors from this data. Also report the dimension of these vectors.
   (b) Write down your personal weight vector (based on your candy preferences) that you will use for computing scores for each of the candies. Annotate your vector with short explanations of what each of the numbers indicate. Your weight vector must have *at least* four non-zero components, and it must have *at least* one component that is not equal to 0 or 1.

(c) What score would your system give to a candy that contains chocolate, is fruit flavored, and contains nougat (but none of the other categories)?

(d) Use R to compute your score for each candy in the dataset and store it as a new column in the dataset. How can you see this as a linear combination of vectors?

(e) Organize the candy by the new score column. What is the highest score your system gives, and for what candy? What about the lowest score?

(f) What are the top-10 candies according to your ranking system? (Did this give you any new recommendations to try?) What are the top-10 candies according to the public's internet voting? How do they compare?

(g) Upload your R script file.

Dataset information: For binary variables, 1 means yes, 0 means no. The data contains the following fields:

- chocolate: Does it contain chocolate?
- fruity: Is it fruit flavored?
- caramel: Is there caramel in the candy?
- peanutalmondy: Does it contain peanuts, peanut butter or almonds?
- nougat: Does it contain nougat?
- crispedricewafer: Does it contain crisped rice, wafers, or a cookie component?
- hard: Is it a hard candy?
- bar: Is it a candy bar?
- pluribus: Is it one of many candies in a bag or box?
- sugarpercent: The percentile of sugar it falls under within the data set.
- pricepercent: The unit price percentile compared to the rest of the set.
- winpercent: The overall win percentage according to 269,000 matchups.