

## Matrices, Music, and Netflix Recommendations

So, the idea is that we can represent data in really useful ways. One way of representing data is called matrices. A matrix is an array of data that has rows and columns.

Ex: The following matrix is data from how five people rank certain movies. The people represent columns, and their names are Abigail, DeMarcus, Julia, Rocky, and Zim. The movies represent rows and are Babadook, Mad Max: Fury Road, Short Term 12, What Happened Miss Simone, and You've Got Mail. Note the movies and people are in alphabetical order for convenience. This matrix's name is the Movie Ratings Matrix, or MRM for short.

$$\begin{pmatrix} 32.0273 & -23.9777 & -7.9933 & 31.9677 & -32.0235 \\ -23.9777 & 43.0225 & -28.9919 & -14.0279 & 23.9755 \\ -7.9933 & -28.9919 & 51.0033 & -22.0087 & 7.9911 \\ 31.9677 & -14.0279 & -22.0087 & 36.0393 & -31.9699 \\ -32.0235 & 23.9755 & 7.9911 & -31.9699 & 32.0273 \end{pmatrix}$$

For example, the first column would represent Abigail's preferences for movies, and she ranks as follows

Babadook: 32.0273  
 Mad Max: Fury Road: -23.9777  
 Short Term 12: -7.9933  
 What Happened Miss Simone: 31.9677  
 You've Got Mail: -32.0235

One useful representation of matrices is called the Singular Value Decomposition (SVD). The idea is to represent the matrix we care about as the sum of matrices that are easily manipulated. Here is an example of a simple matrix. It is a matrix that is represented as a vector multiplied by another vector. So we only need 5 entries to represent 25 entries.

Example for simple matrices:

$$\begin{pmatrix} -13 \\ -12 \\ -4 \\ 16 \\ 13 \end{pmatrix} * \begin{pmatrix} -13 & -12 & -4 & 16 & 13 \end{pmatrix} = \begin{pmatrix} 169 & 156 & 52 & -208 & -169 \\ 156 & 144 & 48 & -192 & -156 \\ 52 & 48 & 16 & -64 & -52 \\ -208 & -192 & -64 & 256 & 208 \\ -169 & -156 & -52 & 208 & 169 \end{pmatrix}$$

How did this multiplication work? Note that the 2nd column and 3rd row of the matrix is 48. In order to get that, we take the 2nd part of the 1st vector, which is -4, and we take the 3rd element of the second vector, which is -12, and multiply them to get 48.

So in order to get a member of the matrix, you multiply the corresponding things in the vectors.

Mathematicians have developed methods of representing matrices in terms of simple matrices. That is the purpose of the SVD. Here is an example

$$\begin{pmatrix} 32.0273 & -23.9777 & -7.9933 & 31.9677 & -32.0235 \\ -23.9777 & 43.0225 & -28.9919 & -14.0279 & 23.9755 \\ -7.9933 & -28.9919 & 51.0033 & -22.0087 & 7.9911 \\ 31.9677 & -14.0279 & -22.0087 & 36.0393 & -31.9699 \\ -32.0235 & 23.9755 & 7.9911 & -31.9699 & 32.0273 \end{pmatrix}$$

The above Matrix is MRM from the first page. The following matrices are A, B, C, D, and E respectively.

A: 
$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

B: 
$$\begin{pmatrix} 16 & -12 & -4 & 16 & -16 \\ -12 & 9 & 3 & -12 & 12 \\ -4 & 3 & 1 & -4 & 4 \\ 16 & -12 & -4 & 16 & -16 \\ -16 & 12 & 4 & -16 & 16 \end{pmatrix}$$

C: 
$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 25 & -35 & 10 & 0 \\ 0 & -35 & 49 & -14 & 0 \\ 0 & 10 & -14 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

D: 
$$\begin{pmatrix} 9 & -6 & -6 & -6 & 9 \\ -6 & 4 & 4 & 4 & -6 \\ -6 & 4 & 4 & 4 & -6 \\ -6 & 4 & 4 & 4 & -6 \\ 9 & -6 & -6 & -6 & 9 \end{pmatrix}$$

E: 
$$\begin{pmatrix} 169 & 156 & 52 & -208 & -169 \\ 156 & 144 & 48 & -192 & -156 \\ 52 & 48 & 16 & -64 & -52 \\ -208 & -192 & -64 & 256 & 208 \\ -169 & -156 & -52 & 208 & 169 \end{pmatrix}$$

As it turns out, (through some voodoo mathematics)  
 $MRM = .0001A + 2B + C + .0002D + .00015E$

So MRM is 2 parts of Matrix B, plus 1 part of Matrix C, plus very small amounts of A, D, and E. If Matrices B and C are the main parts of the matrix, then it would make sense if we only used Matrices B and C, and completely cut out A, D, and E.

2B + C =

$$\begin{pmatrix} 32 & -24 & -8 & 32 & -32 \\ -24 & 43 & -29 & -14 & 24 \\ -8 & -29 & 51 & -22 & 8 \\ 32 & -14 & -22 & 36 & -32 \\ -32 & 24 & 8 & -32 & 32 \end{pmatrix}$$

For comparison, MRM is

$$\begin{pmatrix} 32.0273 & -23.9777 & -7.9933 & 31.9677 & -32.0235 \\ -23.9777 & 43.0225 & -28.9919 & -14.0279 & 23.9755 \\ -7.9933 & -28.9919 & 51.0033 & -22.0087 & 7.9911 \\ 31.9677 & -14.0279 & -22.0087 & 36.0393 & -31.9699 \\ -32.0235 & 23.9755 & 7.9911 & -31.9699 & 32.0273 \end{pmatrix}$$

Note that the matrices are very close to each other. Since the matrices are so close to each other, it would be fair only use 2B + C. And here is the saving in time and energy. B and C are simple matrices. So they only need 5 numbers to represent each. Hence 2B + C only requires 10 numbers to represent it, the 5 numbers for B and the 5 numbers for C. As opposed to MRM which requires 25 numbers to represent it. So we went from 5<sup>2</sup> entries, to 2\*5 entries.

Here's some intuition for what B and C represent. B and C also represent columns of people and rows of movies. In terms of MRM, B and C represent two "ideal data sets" that are a majority of the information. This is related to the ideal people we were talking about in class. Why would we want to only use simple matrices though?

As it turns out, Netflix and Pandora often deal with matrices the size of 100,000<sup>2</sup>. It's a lot of data. A lot of the time, "ideal data sets" come out and people realize the data is well approximated by around 20 or so of these ideal data sets. So that goes from 100,000<sup>2</sup> to 20\*100,000, saving Netflix about 5,000 times as many steps.

What does this mean practically? Imagine Netflix wants to use an algorithm to recommend you some movie selections. This algorithm should be able to run in less than a few seconds so you don't have to wait too long. If the algorithm took 3 seconds to run on 20\*100,000 (the simple data set), then it could potentially take 4 hours to run on the entire data set, 100,000<sup>2</sup>. That's because 3 seconds times 5,000 is a little over 4 hours. So this SVD method takes some algorithms from infeasible, to easily used.

Thanks so much for reading. Lecturing y'all was a blast.