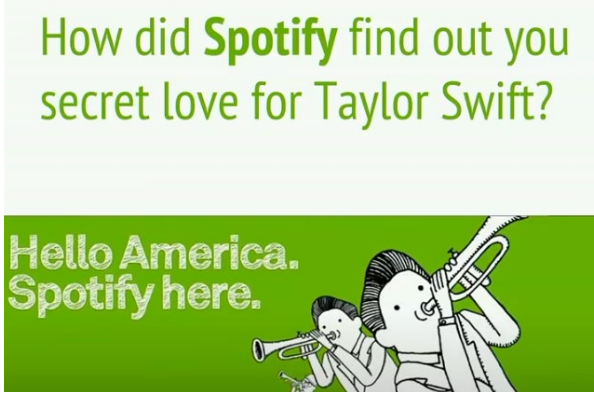
CS 315 - Day 29, Matrix Transformations and SVD



When you are shopping online, how are the recommendations for related items generated? How do Spotify, Netflix, and other apps know what to recommend?

One important element for some recommender systems is the Singular Value Decomposition (SVD). Today we will work toward understanding what the SVD is, and then next class we will explore applications, including image compression.

First, SVDs are based on ***decomposing*** a matrix, which you have seen before in a previous mathematical life.

To spur your memory, let's assume you stopped by the "To-Go" Café before class, and purchased two of the same candy bars and one bottled drink, and your charge was $7.00.

Then, you returned the drink but purchased three more of the same candy bars, and your charge was $4.25.

How much do the candy bars and bottled drinks cost?

There are two variables, c for the cost of the candy bar, and d for the cost of the drink. There were two transactions, which leads to two equations. This is a "classic" two equations – two unknowns type of problem.

$$2c + 1d = 7.00$$

$$3c - 1d = 4.25$$

There are many ways to solve this. One straightforward approach, which happens to work here, is to simply add the two equations. This causes d to cancel out, making c easy to solve for. Once c is calculated, it can be substituted back into one of the equations and d can be solved for:

$5c = 11.25$, $c = 2.25$

$2 \cdot 2.25 + d = 7.00 \rightarrow d = 7.00 - 4.50 \rightarrow d = 2.50$

So, the bottled drink costs $2.50, and the candy bar costs $2.25.

But on larger problems, this approach isn't viable. We could consider this system as a Matrix-Vector product:

$$\begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} 7.00 \\ 4.25 \end{pmatrix}$$

The matrix of coefficients is represented as A, the vector of unknowns is $\bar{x}$, and the right-hand side is b, which gives the linear algebra equation:

$$A\bar{x} = b$$

Let's try Gaussian Elimination on A, where we perform a series of row operations to transform A into an upper-triangular matrix.

$$\begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix}$$

$-\frac{3}{2}R_2 + R_1 \to R_2$ gives $\begin{pmatrix} 2 & 1 \\ 0 & -\frac{5}{2} \end{pmatrix}$

If we keep track of the factors we use to transform the matrix, we get a lower, L and upper, U matrix:

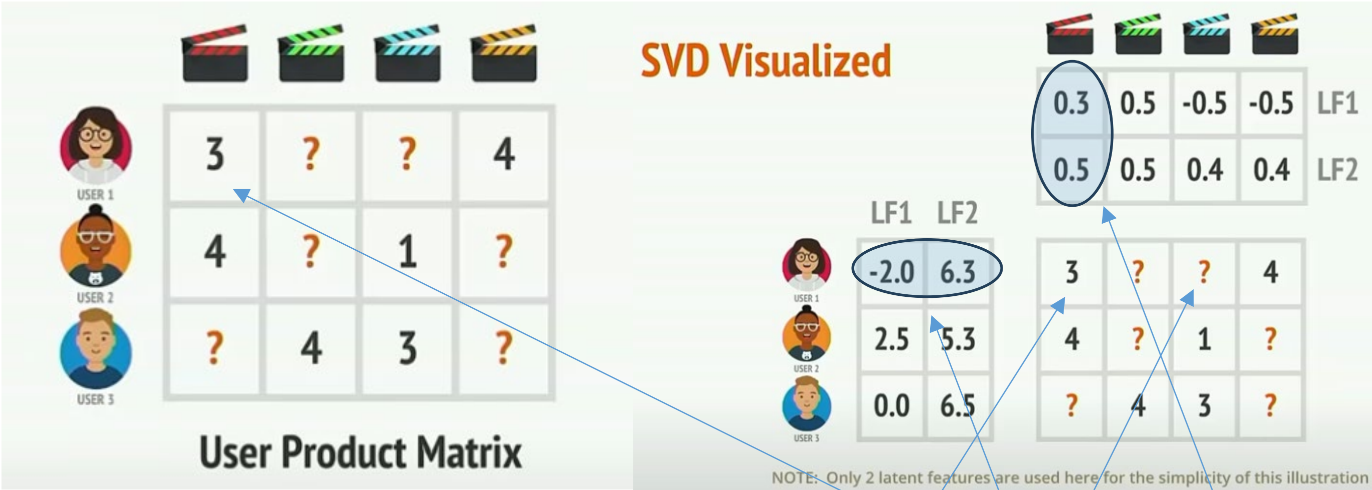
$$A = LU$$

$$\begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{3}{2} & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 0 & -\frac{5}{2} \end{pmatrix}$$

Note that the "lower" indicates the non-zero entries are on the diagonal or below it, and "upper" means the non-zero entries are on the diagonal or above it. And note that given LU we can multiply them together and recreate A. So, this is one of ***many*** ways to decompose a matrix.

The matrix decomposition we'll focus on today is the SVD. And the matrices we'll deal with might be 2D images or based on user rankings:

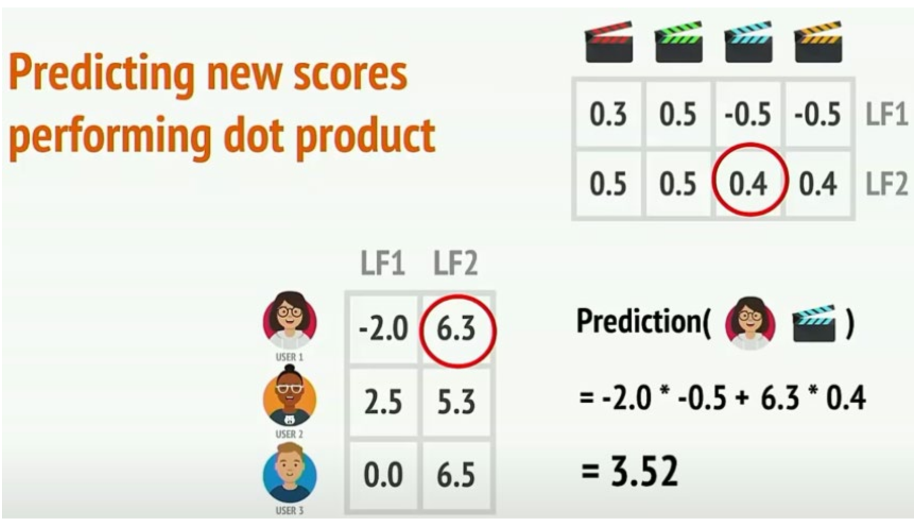
So, what is an SVD? If this is the original User Product Matrix, we could decompose using only a few features (two in the example below). Note: we'll discuss how the latent features are calculated, right now, we're just demonstrating what we can do once we have them. The more features we use, the more accurate our decomposition would be (but it also would be more work).



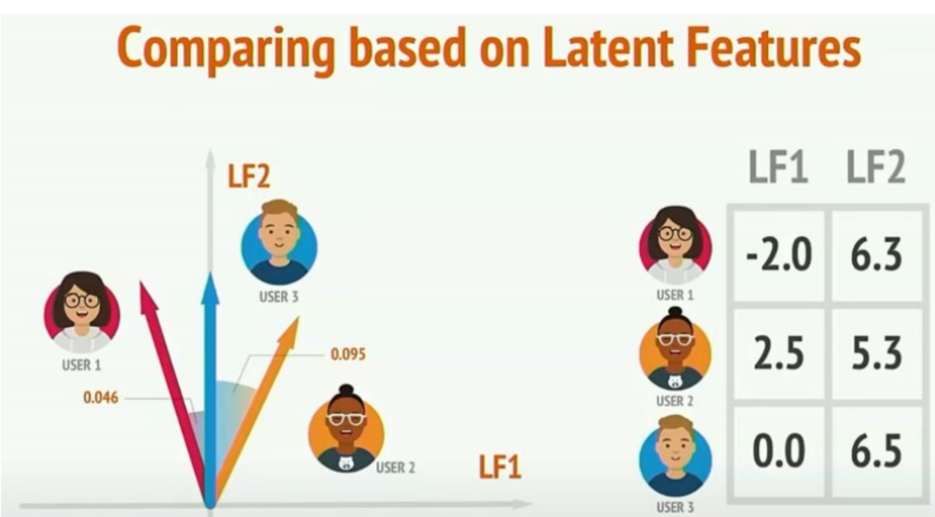
The resulting example will be an approximation. For example, User 1 has a 3 in the top left corner.

The approximation to it, using just the two latent features, is the dot product of the first row and the first column, or $-2.0 \cdot 0.3 + 6.3 \cdot 0.5 = 2.55$, as compared to the original value of 3.

But then we can use this dot product idea to ***predict*** how User 1 might like the "Blue Movie."



That 3.52 suggests that User 1 might like the Blue Movie.



We can also cluster users, by how similar their vectors are. Note that with just two latent features, they give the (x, y) coordinates for a 2D vector.

There are ***many different ways*** to measure similarity between vectors, the one here is cosine similarity, based on the angle between the vectors.

Here, User 1 and User 3 are more similar than User 1 and User 2. So this gives us an idea of how a recommending system may group users to make recommendations.

So, now that we have an idea of how SVD might be used, how do we calculate it and determine those latent feature values?

We need to cover a bit of Linear Algebra, particularly matrix multiplication.

$$B = \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}, C = \begin{pmatrix} 6 & 5 \\ 8 & 9 \end{pmatrix}$$

The product CD is traditionally presented as a dot-product of the rows of C with the columns of D.

$$BC = \begin{pmatrix} 1 \cdot 6 + 4 \cdot 8 & 1 \cdot 5 + 4 \cdot 9 \\ 3 \cdot 6 + 2 \cdot 8 & 3 \cdot 5 + 2 \cdot 9 \end{pmatrix} = \begin{pmatrix} 38 & 41 \\ 34 & 33 \end{pmatrix}$$

<u>In-class w/***assigned groups of four (4)***, write 0. and 1. on board as ground rule(s):</u>

0. No devices.
1. Intro: Name, POE, Icebreaker.
2. Pick someone on the group who will report out.
3. Find CB. Is it the same as BC?
4. Based on this, what can you conclude about the ***commutativity*** (look it up if you don't know the definition) of matrix-matrix multiplication?

$$CB = \begin{pmatrix} 6 \cdot 1 + 5 \cdot 3 & 6 \cdot 4 + 5 \cdot 2 \\ 8 \cdot 1 + 9 \cdot 3 & 8 \cdot 4 + 9 \cdot 2 \end{pmatrix} = \begin{pmatrix} 21 & 34 \\ 35 & 50 \end{pmatrix}$$

5. In the Khan Academy video (https://www.khanacademy.org/math/precalculus/x9e81a4f98389efdf:matrices/x9e81a4f98389efdf:multiplying-matrices-by-matrices/v/multiplying-a-matrix-by-a-matrix) for matrix-matrix multiplication, the matrices $E = \begin{pmatrix} 0 & 3 & 5 \\ 5 & 5 & 2 \end{pmatrix}$ and $D = \begin{pmatrix} 3 & 4 \\ 3 & -2 \\ 4 & -2 \end{pmatrix}$ are used. What shape ED be, that is how many rows and columns? Think rows of E dotted with columns of D. Calculate ED.

$$ED = \begin{pmatrix} 29 & -16 \\ 38 & 6 \end{pmatrix}$$

6. Calculate DE. Note the different shape. Another example demonstrating matrix-matrix multiplication is not commutative.

$$DE = \begin{pmatrix} 20 & 29 & 23 \\ -10 & -1 & 11 \\ -10 & 2 & 16 \end{pmatrix}$$

Back to seats.

For the SVD, we have another way to consider matrix-matrix multiplication, outer products. Basically we will take slices of each matrix, column vectors or row vectors..

Remember

$$B = \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}, C = \begin{pmatrix} 6 & 5 \\ 8 & 9 \end{pmatrix}$$

If we break B into two column vectors, $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} 4 \\ 2 \end{pmatrix}$, and C into two row vectors, $(6 \quad 5)$ and $(8 \quad 9)$, we can calculate BC using outer products:

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} (6 \quad 5) \quad + \quad \begin{pmatrix} 4 \\ 2 \end{pmatrix} (8 \quad 9)$$

$$= \begin{pmatrix} 1 \cdot 6 & 1 \cdot 5 \\ 3 \cdot 6 & 3 \cdot 5 \end{pmatrix} \quad + \quad \begin{pmatrix} 4 \cdot 8 & 4 \cdot 9 \\ 2 \cdot 8 & 2 \cdot 9 \end{pmatrix}$$

$$= \begin{pmatrix} 6 & 5 \\ 18 & 15 \end{pmatrix} \quad + \quad \begin{pmatrix} 32 & 36 \\ 16 & 18 \end{pmatrix}$$

$$= \begin{pmatrix} 38 & 41 \\ 34 & 33 \end{pmatrix}$$

Note, the exact same calculations happen as above, just in different order, which still leads to the same result.

For more info, this module is based on:

- https://www.youtube.com/watch?v=d7iIb_XVkZs
- https://www.tandfonline.com/doi/full/10.1080/07468342.2023.2201567
- https://www.khanacademy.org/math/precalculus/x9e81a4f98389efdf:matrices/x9e81a4f98389efdf:multiplying-matrices-by-matrices/v/multiplying-a-matrix-by-a-matrix
- https://www.3blue1brown.com/lessons/vectors
- https://www.youtube.com/watch?v=kYB8IZa5AuE