Singular Value Decomposition

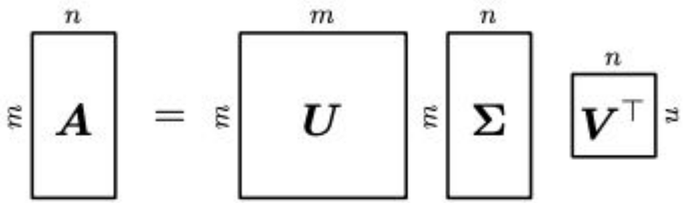
Jayson Cunanan

Simple ideas to begin with

- Dot product is one mathematical way of measuring similarities of '*normalized*' vectors.
  - If x_1 *dot* x_0 > x_2 *dot* x_0 then you can say, x_1 is more similar to x_0 than x_2.
  - Geometrically, this means x_1 points closer to the direction x_0 is pointing.
- Matrix multiplication is just computing similarities set-wise.
  - Think of **A** as a collection of row vectors and **B** as column vectors. Then **A** *matmul* **B** gives the similarities of the rows of **A** with the columns of **B**.
  - If you are comparing rows for, say, **P** and **Q**, then take the transpose of the second matrix to convert rows to columns. I.e. **P** *matmul* $\mathbf{Q}^T$
- Rank of a matrix is the maximum number of linearly independent rows/cols

SVD Theorem

Theorem 4.22 (SVD Theorem). *Let* $A^{m \times n}$ *be a rectangular matrix of rank* $r \in [0, \min(m, n)]$. *The SVD of* $\boldsymbol{A}$ *is a decomposition of the form*



(4.64)

with an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ *with column vectors* $\boldsymbol{u}_i$, $i = 1, \ldots, m$, *and an orthogonal matrix* $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ *with column vectors* $\boldsymbol{v}_j$, $j = 1, \ldots, n$. *Moreover,* $\boldsymbol{\Sigma}$ *is an* $m \times n$ *matrix with* $\Sigma_{ii} = \sigma_i \geqslant 0$ *and* $\Sigma_{ij} = 0,\ i \neq j$.

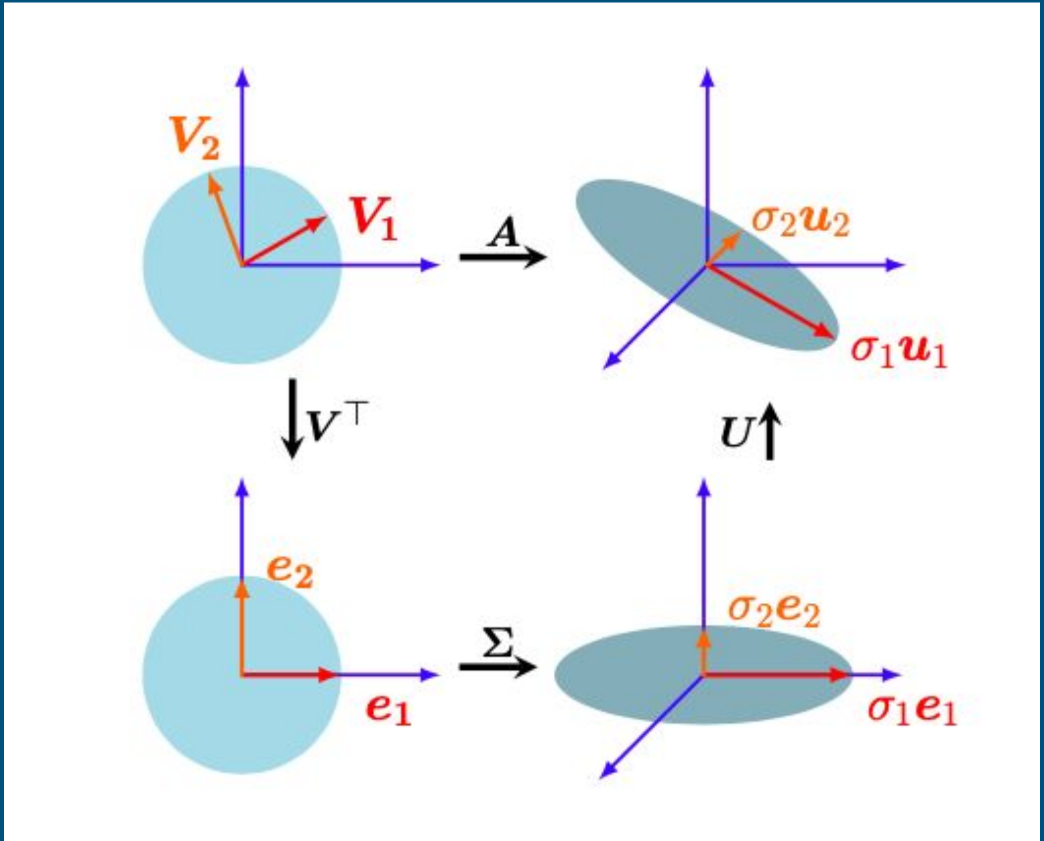
Key points

- SVD applies to all matrices and always exists.
- Orthogonal matrices **U** and **V** act as basis changes.
- Singular values are arranged in decreasing order,.
    - i.e. /sigma_1 >=/sigma_2>= . . . /sigma_r
- If ∑ is rectangular, it would have zero padding.

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}.$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & & 0 \\ 0 & 0 & \sigma_m & 0 & \dots & 0 \end{bmatrix}$$

Geometric intuition



SVD applies a sequence of linear transformations

1. **Rotation (domain)**
2. **Stretching/Shrinking w/ the possibility of adding a new dimension**
3. **Rotation (codomain)**

Eigenvalue Decomposition vs. Singular Value Decomposition

$A = PDP^{-1}$	$A = U\Sigma V^T$
Only for square matrices and exists only if P exists	Always exist
Basis change not necessarily rotations	basis change are rotations
Composed of 3 linear mappings	Composed of 3 linear mappings
Domain and codomain have same dimensions	Domain and codomain can be different dimensions
Diagonal entries are NOT always real and nonnegative	Diagonal entries are always real and nonnegative

Example 4.14 (Finding Structure in Movie Ratings and Consumers)

Sci-fi?	French?	Love?	War?

$$
\begin{array}{r c}
 & \begin{array}{ccc} \text{Ali} & \text{Beatrix} & \text{Chandra} \end{array} \\
\begin{array}{r} \text{Star Wars} \\ \text{Blade Runner} \\ \text{Amelie} \\ \text{Delicatessen} \end{array} &
\begin{bmatrix} 5 & 4 & 1 \\ 5 & 5 & 0 \\ 0 & 0 & 5 \\ 1 & 0 & 4 \end{bmatrix}
\end{array}
=
\begin{bmatrix}
-0.6710 & 0.0236 & 0.4647 & -0.5774 \\
-0.7197 & 0.2054 & -0.4759 & 0.4619 \\
-0.0939 & -0.7705 & -0.5268 & -0.3464 \\
-0.1515 & -0.6030 & 0.5293 & -0.5774
\end{bmatrix}
\begin{bmatrix}
9.6438 & 0 & 0 \\
0 & 6.3639 & 0 \\
0 & 0 & 0.7056 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
-0.7367 & -0.6515 & -0.1811 \\
0.0852 & 0.1762 & -0.9807 \\
0.6708 & -0.7379 & -0.0743
\end{bmatrix}
$$

Space of stereotypical movies (row entries)

Space of stereotypical viewers (column entries)

Matrix approximation

- Observation:

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i \ ,$$

- Action/Try

$$\widehat{\boldsymbol{A}}(k) := \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{k} \sigma_i \boldsymbol{A}_i$$