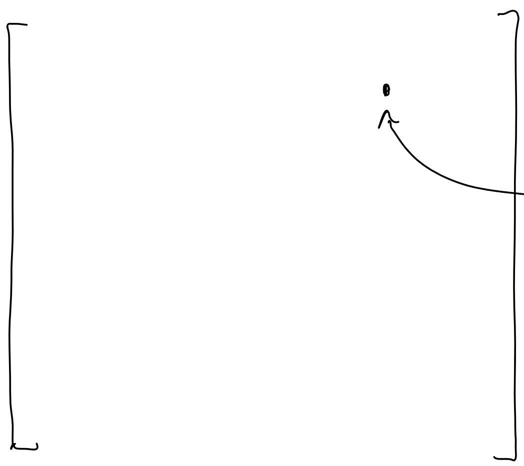
Intro: SVD

Decomposing a matrix. Important and common.

<u>Motivation</u>: "Netflix Challenge problem", equivalent to a "matrix completion" problem.

Predict which users will like certain movies.

Giant Matrix: rows = people / netflix subscribers.
Columns = movies.
(i,j) = the rating ~~of~~ the movie, given by person i to movie j.



row i, column j. $\Big\}$ the rating.

some entries of this matrix = unknown

<u>Goal</u>: Replace '?' with numbers that represent true preferences,

High level:

$$\begin{pmatrix} 1 & \cancel{?}^{1} & \cancel{?}^{1.5} \\ \cancel{?}^{2} & 2 & \cancel{?}^{3} \\ \cancel{?}^{6} & 6 & 9 \\ \cancel{?}^{2} & \cancel{?}^{2} & 3 \\ 4 & 4 & \cancel{?}^{6} \end{pmatrix}$$

Impossible if no additional information. But if give more info:

Additional information: each row is a multiple of other rows.

$\equiv$ matrix has rank 1.

Rank-0 matrix $\equiv$ all zeros matrix.

Rank-1 matrix $\equiv$ all rows or columns are multiples of each other.

Equivalently, if we have a rank-1 matrix,

$\underset{\substack{m \times n \\ \text{matrix}}}{A} = \underbrace{u \cdot v^T}_{\text{outer product}}$, $\underset{\substack{m \times 1 \\ \text{vector}}}{u}$ are $\underset{\substack{n \times 1 \\ \text{vector}}}{v}$ are vectors

ij^{th} entry of $A = u_i \cdot v_j$

Note: outer product: $(m \times 1)(1 \times n) = m \times n$.
inner product: $(m \times 1)(1 \times n) = 1$

$$A = u \cdot v^T = \begin{array}{cc} & 1 \\ m & \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \end{array} \begin{array}{cc} n & \\ \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} & 1 \end{array}$$

$$= \begin{array}{cc} m & \begin{bmatrix} u_1 v_1 & \cdots & u_1 v_n \\ \vdots & \ddots & \\ u_m v_1 & & u_m v_n \end{bmatrix} \\ & n \end{array}$$

$$= \begin{bmatrix} u_1 \cdot v^T \\ u_2 \cdot v^T \\ \vdots \\ u_m \cdot v^T \end{bmatrix} = \begin{array}{cc} \left[\begin{array}{c|c|c|c} v_1 \cdot u & v_2 \cdot u & \cdots & v_n \cdot u \\ \vdots & & & \end{array}\right] & m \\ n & \end{array}$$

This is the case of a rank-1 matrix.

Now consider case where A is a rank-2 matrix. This means A is the sum of two rank-1 matrices (and A is not rank-1).

$$\overset{\text{rank 2}}{A} = \underbrace{u \cdot v^T}_{\text{rank-1}} + \underbrace{w z^T}_{\text{rank-1}} =$$

$$\underbrace{\begin{bmatrix} u_1 v^T + w_1 \cdot z^T \\ \vdots \\ u_m v^T + w_m \cdot z^T \end{bmatrix}}_{\text{row form}} = \underbrace{\begin{bmatrix} v_1 \cdot u + z_1 \cdot w & \cdots & v_n \cdot u + z_n \cdot w \end{bmatrix}}_{\text{column form}}$$

$$= \underbrace{m\Bigg\{\underbrace{\begin{bmatrix} | & | \\ u & w \\ | & | \end{bmatrix}}_{2} \; 2\Big\{\underbrace{\begin{bmatrix} — & v^T & — \\ — & z^2 & — \end{bmatrix}}_{n}}_{\text{form : outer product of vectors}} = m\Bigg\{\underbrace{\begin{bmatrix} \bigcirc & & \\ & & \\ u_1 v_1 + w_1 z_1 & & \end{bmatrix}}_{n}$$

Define the SVD of a matrix:

orthogonal: $XX^T = X^TX = I$, transpose is its inverse.

Every matrix $A = U \cdot S \cdot V^T$

A: m by n

U: $m \times m$ orthogonal matrix

S: $m \times n$ diagonal matrix

V^T: $n \times n$ orthogonal matrix. rows of V^T / columns of V are "right singular vectors".

Columns of U are left singular vectors.

Entries of S: $s_1 \geq s_2 \geq \cdots \geq 0$.

s = singular values.

$$\Downarrow$$

$$\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_n \\ \hline & & 0 & \end{bmatrix} \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}^T$$

eigenvector; orthonormal = orthogonal + unit length.

Each column = basis of corresponds column in A.

u_1 more important than $u_2, u_3, \ldots$

Likewise s_1 more important than $s_2 \ldots$

Likewise, v_1 more important than $v_2, \ldots$

"Interpretation":
- a_i = some description (face).
- u_i = normalized version of a_i
- s_i = importance of "features"
- v_i^T = mixture of "features"

Then,
we can then select only first k columns of U and V, and first k values of S, that approximates A.

$$\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{bmatrix} \left[\begin{array}{cccc} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_n \\ \hline & & 0 & \end{array}\right]_{\uparrow \text{ or } m} \begin{bmatrix} - \; v_1^T \; - \\ - \; v_2^T \; - \\ \vdots \\ - \; v_n^T \; - \end{bmatrix}$$

$$A = s_1 u_1 v_1^T + s_2 u_2 v_2^T + \cdots + s_n u_n v_n^T + 0 .$$

$$= s_1 \underset{u_1}{\Big|}\overset{v_1^T}{\boxed{\text{rank 1 matrix}}} + s_2 \underset{u_2}{\Big|}\overset{v_2^T}{\boxed{\text{rank 1 matrix}}} + \cdots + \underset{\uparrow \text{ or } m}{s_n} \underset{u_n}{\Big|}\overset{v_n^T}{\boxed{\text{rank 1 matrix}}}$$

$$A = \sum_{i=1}^{\min(n,m)} s_i u_i \otimes v_i^T$$

outer product.

truncate at rank k, since most information captured in the first k matrix.

The singular values are unique.

The singular vectors are not unique.

SVD can be computed in time $O(m^2 n)$ or $O(n^2 m)$, whichever is smaller.

we could just <u>define</u> a matrix to have rank-k

if

$$m\overset{n}{\begin{bmatrix} A \end{bmatrix}} = m\overset{k}{\begin{bmatrix} Y \end{bmatrix}} \, k\overset{n}{\begin{bmatrix} Z^T \end{bmatrix}} \quad (\text{and } A \text{ is not rank } 0, \cdots, k-1)$$

Given matrix A, it would be great if we could find a factorization $Y \cdot Z^T$ where k is really small.

A $m \cdot n$ entries can be written using only $k(m+n)$ numbers.

one way we could represent A (approximation with rank k):

$A = USV^T$

$$
\begin{array}{cc} & n \\ m & \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \\ & & & 0 \end{pmatrix} \end{array}
$$

zero out some entries of S (with smallest entries),

S $m \times n$ matrix $\longrightarrow$
$$
\begin{array}{cc} & k \\ k & \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_k \end{pmatrix} \end{array}
$$
diagonal

$$
\underset{\substack{m \times n, \\ \text{still.}}}{A} = \underset{m \times k}{U} \cdot \underset{k \times k}{S} \cdot \underset{k \times n}{V^T}
$$

$m \times k$: taken only k columns from U.

$k \times n$: k rows from V^T.

<u>Define Frobenius Norm</u>: $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$

<u>Goal</u>:

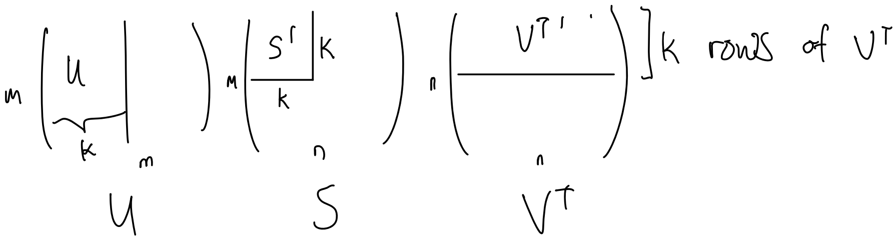
Given a matrix A, goal is to find matrix A'
s.t. A' has rank k and minimizes
$\|A - A'\|_F$ over all rank k matrices.

Answer: compute SVD of A and take the top k
singular vectors and values,

(Eckard-Young Theorem [1936]: $\underset{\tilde{X} \text{ s.t. } \text{rank}(\tilde{X}) = k}{\text{argmin}} \left(\|X - \tilde{X}\|_F \right) = \tilde{U}\tilde{S}\tilde{V}^T$).

$A = USV^T$.

Answer is $A' = U'S'V^{T'}$ (here, $'$ = $\sim$).



A' is still $m \times n$. Note, after truncation:

$\tilde{U}^T \tilde{U} = I_{k \times k}$ but $\tilde{U}\tilde{U}^T \neq I$

Application to matrix completion

Matrix Completion: A

- Replace '?' with either:
  - 0
  - Average value of known entries
  - Average value in that column or row.
- Find the best rank k approximation to A after filling in the '?'s.
- output this best rank k approximation.

How to choose k? k is a hyperparameter.

One typical heuristic for choosing k is to take enough singular values so that the sum of the remaining values $\leq \frac{1}{10}$ of values you did take.

Application: Linear Regression

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$$

A: $m \times n$ matrix; $b \in \mathbb{R}^m$ (vector of length m).

Easy case: $A = D$ (diagonal).

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}$$

Want: $x_1 = \frac{b_1}{d_1}$ $\quad x_2 = \frac{b_2}{d_2}$ $\quad d_j = 0 \Rightarrow x_j = 0$

Solution is

$$\begin{array}{c} \\ m \end{array} \underbrace{\begin{pmatrix} \frac{1}{d_1} & & & \\ & \frac{1}{d_2} & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}}_{n} \cdot \begin{array}{c} 1 \\ \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}}_{b = \text{vector of } m \times 1} \end{array} m = m \begin{array}{c} m \\ \begin{bmatrix} \frac{b_1}{d_1} & \frac{b_2}{d_2} & \cdots \end{bmatrix} \end{array}$$

D^+ = "pseudo-inverse" of D.

Summarize: $\boxed{X = D^+ \cdot b}$, since $\overset{A}{D}\overbrace{D^+ b}^{X} - b = b - b = 0$.

$\underbrace{DD^+}_{I}$

0 ← minimized.

General case (A is non-square).

$$\|Ax-b\|^2 \equiv \min_x \|\underbrace{USV^T}_{A}x - b\|^2$$

Note: $\|Ux\| = \|x\|$

Orthogonal matrix, with norm 1.

multiplied by U^T, which has norm 1, so equivalent step.

$$\equiv \min_x \|SU^Tx - U^Tb\|^2$$

substitute $y = V^Tx \Rightarrow Vy = VU^Tx = x$

$$= \min_x \|Sy - U^Tb\|^2$$

diagonal

$$\Rightarrow S^+Sy - S^+U^Tb \text{ , set } = 0$$

$y = V^Tx$

$$\Rightarrow y = S^+ \cdot U^T b$$

$$\Rightarrow V^T x = S^+ \cdot U^T b$$

multiply by V

$$\Rightarrow \boxed{x = VS^+U^Tb}$$

solution to linear least squares, using SVD.

Example: PCA : Find Eigendecomposition of covariance matrix.

$$A = X^T X = (\underbrace{USV^T}_{X})^T \cdot \underbrace{USV^T}_{X}$$

$$= VSU^T \cdot \underbrace{U^T U}_{I} \cdot SV^T$$

$$= VS^2V^T$$

eigendecomposition completed!

ortho ortho

right singular vectors of X (rows of V^T) are the principal components (top eigenvectors of X^TX).

singular values are the square root of eigenvalues of X^TX. Square the singular values to get eigenvalues.

One Further application: Image compression

Consider a Black and white image.

matrix A $m \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}$, each entry is $\{0,1\}$.

n

For compression, Compute A' = low-rank approximation of A for some value k.

entries of A' are numbers btwn $0, 1$, and not $\{0,1\}$.