Singular-Value Decomposition with Applications

Anil Maheshwari

anil@scs.carleton.ca
School of Computer Science
Carleton University
Canada

Outline

Matrices

Eigenvalues and Eigenvectors

Symmetric Matrices

Singular Value Decomposition

Applications

CUR

Matrices

Matrices

1. A Rectangular Array
2. Operations: Addition; Multiplication; Diagonalization; Transpose; Inverse; Determinant
3. Row Operations; Linear Equations; Gaussian Elimination
4. Types: Identity; Symmetric; Diagonal; Upper/Lower Traingular; Orthogonal; Orthonormal
5. Transformations - Eigenvalues and Eigenvectors
6. Rank; Column and Row Space; Null Space
7. Applications: Page Rank, Dimensionality Reduction, Recommender Systems, . . .

Utility Matrix M

A Matrix M where rows represent users, columns items, and entries in M represents the ratings.

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} .13 & -.02 & .01 \\ .41 & -.07 & .03 \\ .55 & -.1 & .04 \\ .68 & -.11 & .05 \\ .15 & .59 & -.65 \\ .07 & .73 & .67 \\ .07 & .29 & -.32 \end{bmatrix} \begin{bmatrix} 12.5 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.35 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \\ .40 & -.8 & .40 & .09 & .09 \end{bmatrix}$$

Questions: How to guess missing entries? How to guess ratings for a new user? . . .

Matrix Vector Product

- Matrix-vector product: $Ax = b$

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

- $Ax = b$ as linear combination of columns:

$$\begin{aligned} \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \end{bmatrix} &= 4 \begin{bmatrix} 2 \\ 3 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 8 \\ 12 \end{bmatrix} \begin{bmatrix} -2 \\ -8 \end{bmatrix} \\ &= \begin{bmatrix} 6 \\ 4 \end{bmatrix} \end{aligned}$$

Matrix-Matrix Product

- Matrix-matrix product $A = BC$:

$$\begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 6 & 16 \end{bmatrix}$$

- $A = BC$ as sum of rank 1 matrices:

$$\begin{aligned} \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 4 \end{bmatrix} &= \begin{bmatrix} 2 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 8 \\ 6 & 12 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 8 \\ 6 & 16 \end{bmatrix} \end{aligned}$$

Row Reduced Echelon Form

Let $A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 4 & 8 \\ 10 & 16 & 24 \end{bmatrix}$

1st Pivot: Replace r_2 by $r_2 - r_1$, and r_3 by $r_3 - 5r_1$:

$$\begin{bmatrix} 2 & 2 & 0 \\ 0 & 2 & 8 \\ 0 & 6 & 24 \end{bmatrix}$$

2nd Pivot: Replace r_3 by $r_3 - 3r_2$:

$$\begin{bmatrix} 2 & 2 & 0 \\ 0 & 2 & 8 \\ 0 & 0 & 0 \end{bmatrix}$$

RREF contd.

Divide the first row by 2, the second row by 2:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

Replace r_1 by $r_1 - r_2$:

$$R = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 4 & 8 \\ 10 & 16 & 24 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} = R$$

Definitions:

- Rank $=$ Number of non-zero pivots $= 2$
- Basis vectors of row space = rows corresponding to non-zero pivots in R
  $v_1 = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$
- Basis vectors of column space = Columns of A corresponding to non-zero pivots of R.

  $u_1 = \begin{bmatrix} 2 \\ 2 \\ 10 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 2 \\ 4 \\ 16 \end{bmatrix}$
- A as sum of the product of rank 1 matrices

$$A = u_1 v_1^T + u_2 v_2^T = \begin{bmatrix} 2 \\ 2 \\ 10 \end{bmatrix} \begin{bmatrix} 1 & 0 & -4 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 16 \end{bmatrix} \begin{bmatrix} 0 & 1 & 4 \end{bmatrix}$$

Null Space

Null space of A = All vectors x such that $Ax = 0$.
This includes the 0 vector $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Is there a vector $x = (x_1, x_2, x_3) \in R^3$, such that
$Ax = x_1 \begin{bmatrix} 2 \\ 2 \\ 10 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \\ 16 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 8 \\ 24 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$x = (1, -1, 1/4)$, or any of its scalar multiples, satisfies $Ax = 0$

Dimension of Null Space of A= Number of columns (A) - rank(A)= $3 - 2 = 1$

Spaces for A

Let A be $m \times n$ matrix with real entries.
Let R be RREF of A consisting of $r \leq \min\{m, n\}$ non-zero pivots.

1. rank$(A) = r$
2. Column space is a subspace of R^m of dimension r, and its basis vectors are the columns of A corresponding to the non-zero pivots in R.
3. Row space is a subspace of R^n of dimension r, and its basis vectors are the rows of R corresponding to the non-zero pivots.
4. The null-space of A consists of all the vectors $x \in R^n$ satisfying $Ax = 0$. They form a subspace of dimension $n - r$.

Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

Given an $n \times n$ matrix A.
A non-zero vector v is an eigenvector of A, if $Av = \lambda v$ for some scalar λ.
λ is the eigenvalue corresponding to vector v.

Example

Let $A = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$

Observe that

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 3 \end{bmatrix} \text{ and } \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Thus, $\lambda_1 = 5$ and $\lambda_2 = 1$ are the eigenvalues of A.
Corresponding eigenvectors are $v_1 = [1, 3]$ and $v_2 = [1, -1]$, as $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$.

Matrices with distinct eigenvalues

Propertry

Let A be an $n \times n$ real matrix with n distinct eigenvalues.
The corresponding eigenvectors are linearly independent.

Proof: Proof by contradiction. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues and $v_1, \ldots, v_n$ the corresponding eigenvectors, that are linearly dependent.

Assume $v_1, \ldots, v_{n-1}$ are L.I. (otherwise work with a smaller set).

Dependence $\implies \alpha_1 v_1 + \ldots + \alpha_{n-1} v_{n-1} + \alpha_n v_n = 0$, where $\alpha_n \neq 0$.

$\implies v_n = \frac{-\alpha_1}{\alpha_n} v_1 + \ldots + \frac{-\alpha_{n-1}}{\alpha_n} v_{n-1}$

Multiply by A: $A v_n = \lambda_n v_n = \frac{-\alpha_1}{\alpha_n} \lambda_1 v_1 + \ldots + \frac{-\alpha_{n-1}}{\alpha_n} \lambda_{n-1} v_{n-1}$

Multiply by λ_n: $\lambda_n v_n = \frac{-\alpha_1}{\alpha_n} \lambda_n v_1 + \ldots + \frac{-\alpha_{n-1}}{\alpha_n} \lambda_n v_{n-1}$

Subtract last two equations:
$0 = \frac{-\alpha_1}{\alpha_n} (\lambda_n - \lambda_1) v_1 + \ldots + \frac{-\alpha_{n-1}}{\alpha_n} (\lambda_n - \lambda_{n-1}) v_{n-1}$

Since, $\lambda_n - \lambda_i \neq 0, \implies$ the vectors $v_1, \ldots, v_{n-1}$ arre linearly dependent.
A contradiction.

$\square$

Matrices with distinct eigenvalues

Let A be an $n \times n$ real matrix with n distinct eigenvalues.
Let $\lambda_1, \dots, \lambda_n$ be the distinct eigenvalues and let $x_1, \dots, x_n$ be the corresponding eigenvectors, respectively. Let each $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]$.

Define an eigenvector matrix $X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{nn} \end{bmatrix}$

Define a diagonal $n \times n$ matrix Λ $\begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \lambda_n \end{bmatrix}$

Consider the matrix product AX,

$$AX = A \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \dots & \lambda_n x_n \end{bmatrix} = X\Lambda$$

Matrices with distinct eigenvalues

Since eigenvectors are linearly independent, we know that X^{-1} exists.

Multiply by X^{-1} on both the sides from left in $AX = X\Lambda$ and we obtain

$$X^{-1}AX = X^{-1}X\Lambda = \Lambda \tag{1}$$

and when we multiply on the right we obtain

$$AXX^{-1} = A = X\Lambda X^{-1} \tag{2}$$

An Application of Diagonalization $A = X\Lambda X^{-1}$

Consider $A^2 = (X\Lambda X^{-1})(X\Lambda X^{-1}) = X\Lambda(X^{-1}X)\Lambda X^{-1} = X\Lambda^2 X^{-1}$
$\implies A^2$ has the same set of eigenvectors as A, but eigenvalues are squared.

Similarly, $A^k = X\Lambda^k X^{-1}$.
Eigenvectors of A^k are same as that of A and its eigenvalues are raised to the power of k.

Symmetric Matrices

Example

Consider symmetric matrix $S = \left[\begin{smallmatrix} 3 & 1 \\ 1 & 3 \end{smallmatrix}\right]$.
Its eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = 2$ and the corresponding eigenvectors are $q_1 = (1/\sqrt{2}, 1/\sqrt{2})$ and $q_2 = (1/\sqrt{2}, -1/\sqrt{2})$, respectively.
Note that eigenvalues are real and the eigenvectors are orthonormal.

$$S = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Eigenvalues of Symmetric Matrices

All the eigenvalues of a real symmetric matrix S are real. Moreover, all components of the eigenvectors of a real symmetric matrix S are real.

Symmetric Matrices (contd.)

Property

Any pair of eigenvectors of a real symmetric matrix S corresponding to two different eigenvalues are orthogonal.

Proof: Let q_1 and q_2 be eigenvectors corresponding to $\lambda_1 \neq \lambda_2$, respectively.
We have $Sq_1 = \lambda_1 q_1$ and $Sq_2 = \lambda_2 q_2$.
Now $(Sq_1)^T = q_1^T S^T = q_1^T S = \lambda_1 q_1^T$, as S is symmetric,
Multiply by q_2 on the right and we obtain $\lambda_1 q_1^T q_2 = q_1^T S q_2 = q_1^T \lambda_2 q_2$.
Since $\lambda_1 \neq \lambda_2$ and $\lambda_1 q_1^T q_2 = q_1^T \lambda_2 q_2$, this implies that $q_1^T q_2 = 0$ and thus the eigenvectors q_1 and q_2 are orthogonal.

$\square$

Symmetric Matrices (contd.)

Symmetric matrices with distinct eigenvalues

Let S be a $n \times n$ symmetric matrix with n distinct eigenvalues and let $q_1, \ldots, q_n$ be the corresponding orthonormal eigenvectors. Let Q be the $n \times n$ matrix consiting of $q_1, \ldots, q_n$ as its columns. Then
$S = Q\Lambda Q^{-1} = Q\Lambda Q^T$.
Furthermore, $S = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T$

An Example:

$$
\begin{aligned}
S = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \\
&= 4 \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} + 2 \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}
\end{aligned}
$$

Theorem

For a real symmetric $n \times n$ matrix S, we have

1. *All eigenvalues of S are real.*
2. *S can be expressed as $S = Q\Lambda Q^T$, where Q consists of orthonormal basis of R^n formed by n eigenvectors of S, and Λ is a diagonal matrix consisting of n eigenvalues of S.*
3. *S can be expressed as the sum of the product of rank 1 matrices:*

$$S = \lambda_1 q_1 q_1^T + \ldots + \lambda_n q_n q_n^T$$

Inverse of Symmetric Matrices

Claim

$S = Q\Lambda Q^T$ and $S^{-1} = \frac{1}{\lambda_1}q_1q_1^T + \ldots + \frac{1}{\lambda_n}q_nq_n^T$

Proof Sketch: $S = Q\Lambda Q^T = \lambda_1 q_1 q_1^T + \ldots + \lambda_n q_n q_n^T$

$SS^{-1} = (\lambda_1 q_1 q_1^T + \ldots + \lambda_n q_n q_n^T)(\frac{1}{\lambda_1}q_1q_1^T + \ldots + \frac{1}{\lambda_n}q_nq_n^T) = I$ as $q_1, \ldots, q_n$ are orthonormal.

$\square$

Positive Definite Matrices

A symmetric matrix S is positive definite if all its eigenvalues > 0.
It is positive semi-definite if all the eigenvalues are ≥ 0.

An Alternate Characterization

Let S be a $n \times n$ real symmetric matrix. For all non-zero vectors $x \in R^n$, if $x^T S x > 0$ holds, then all the eigenvalues of S are > 0.

Proof: Let λ_i be an eigenvalue of S.
Let the corresponding unit eigenvector is q_i.
Note that $q_i^T q_i = 1$.
Since S is symmetric, we know that λ_i is real.
Now we have, $\lambda_i = \lambda_i q_i^T q_i = q_i^T \lambda_i q_i = q_i^T S q_i$.
But $q_i^T S q_i > 0$, hence $\lambda_i > 0$.

$\square$

Singular Value Decomposition

Square Matrices:

A be an $n \times n$ matrix with distinct eigenvalues.

$X_{n \times n}$ = Matrix of eigenvectors of A

$AX = X\Lambda$, $A = X\Lambda X^{-1}$, $\Lambda = X^{-1}\Lambda X$

Symmetric Matrices:

S be an $n \times n$ symmetric matrix with distinct eigenvalues.

$Q_{n \times n}$= Matrix of n-orthonormal eigenvectors of S

$S = Q\Lambda Q^T$

What if A is a rectangular matrix of dimensions $m \times n$?

SVD of Rectangular Matrices

Let A be a $m \times n$ matrix of rank r with real entries.

We can find orthonormal vectors in $\mathbb{R}^n$ such that their product with A results in a scaled copy of orthonormal vectors in $\mathbb{R}^m$.

Formally, we can find

1. Orthonormal vectors $v_1, \ldots, v_r \in \mathbb{R}^n$
2. Orthonormal vectors $u_1, \ldots, u_r \in \mathbb{R}^m$
3. Real numbers $\sigma_1, \ldots, \sigma_r \in \mathbb{R}$
4. For $i = 1, \ldots, r$: $Av_i = \sigma_i u_i$
5. $AV = U\Sigma$, i.e.,

$$A \begin{bmatrix} v_1 & \ldots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & \ldots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \sigma_r \end{bmatrix}$$

6. $A = U\Sigma V^T$

Example

An Example: $AV = U\Sigma$

$$\begin{bmatrix} 1 & 5 \\ 0 & 3 \\ 1 & 4 \\ 4 & 0 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} .60 & -.8 \\ .8 & .6 \end{bmatrix} = \begin{bmatrix} .58 & .39 \\ .31 & .30 \\ .48 & .28 \\ .30 & -.56 \\ .48 & -.59 \end{bmatrix} \begin{bmatrix} 7.8 & 0 \\ 0 & 5.7 \end{bmatrix}$$

Alternatively, $A = U\Sigma V^T$

$$\begin{bmatrix} 1 & 5 \\ 0 & 3 \\ 1 & 4 \\ 4 & 0 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} .58 & .39 \\ .31 & .30 \\ .48 & .28 \\ .30 & -.56 \\ .48 & -.59 \end{bmatrix} \begin{bmatrix} 7.8 & 0 \\ 0 & 5.7 \end{bmatrix} \begin{bmatrix} .60 & .8 \\ -.8 & .6 \end{bmatrix}$$

Play around with the SVD command in Wolfram Alpha for some matrices.

Matrix $A^T A$

Symmetric and Positive semi-definite

Let A be $m \times n$ matrix, where $m \geq n$. The matrix $A^T A$ is symmetric and positive semi-definite

Proof:
Symmetric: $(A^T A)^T = A^T (A^T)^T = A^T A$

Positive semi-definite: Take any non-zero vector $x \in \mathbb{R}^n$
$x^T (A^T A) x = (x^T A^T)(Ax) = (Ax)^T (Ax) = ||Ax||^2 \geq 0$

$\square$

$A^T A$ is a symmetric matrix of dimension $n \times n$. Eigenvalues of $A^T A$ are non-negative and the corresponding eigenvectors are orthonormal.

Let $\lambda_1 \geq \ldots \geq \lambda_n$ be eigenvalues of $A^T A$ and let $v_1, \ldots, v_n$ be the corresponding eigenvectors.

$$A^T A v_i = \lambda_i v_i \Leftrightarrow v_i^T A^T A v_i = \lambda_i$$

Define $\sigma_i = ||Av_i|| \implies \sigma_i^2 = ||Av_i||^2 = v_i^T A^T A v_i = \lambda_i$

Hence, $\sigma_i = ||Av_i|| = \sqrt{\lambda_i}$

Consider two cases:

Full Rank: Rank of $A^T A$ is n.

Low Rank: Rank of $A^T A$ is $r < n$.

Matrix $A^T A$ is Full Rank

Assume, $\sigma_1 \geq \ldots \geq \sigma_n > 0$
($\implies A$ and $A^T A$ has rank n)

Define vectors $u_1, \ldots, u_n \in \mathbb{R}^m$ as $u_i = Av_i/\sigma_i$

Orthonormal

The set of vectors $u_i = Av_i/\sigma_i$, for $i = 1, \ldots, n$, are orthonormal.

Proof: $||u_i|| = ||Av_i||/\sigma_i = \sigma_i/\sigma_i = 1$

Consider the dot product of any two vectors u_i and u_j:
$u_i^T u_j = (Av_i/\sigma_i)^T (Av_j/\sigma_j) = \frac{1}{\sigma_i \sigma_j} v_i^T A^T A v_j = \frac{1}{\sigma_i \sigma_j} v_i^T \lambda_j v_j = \frac{\lambda_j}{\sigma_i \sigma_j} v_i^T v_j = 0$

$\square$

Matrix $A^T A$ is Low Rank

Suppose $m \geq n$, but $rank(A) = r < n$.

Eigenvalues of $A^T A$

The $n - r$ eigenvalues of $A^T A$ are equal to 0.

Proof: Consider a basis of the null space of A.
Let $x_1, \ldots, x_{n-r}$ be a basis of the null space of A.
This implies that $Ax_j = 0$ for $j = 1, \ldots, n - r$.
Now, $A^T Ax_j = 0 = 0x_j$.
Thus, 0 is an eigenvalue of $A^T A$ corresponding to each x_i's.
Thus $n - r$ eigenvalues of $A^T A$ are equal to 0

$\square$

Handling low rank (contd.)

Consider eigenvalues and eigenvectors of $A^T A$
Let $\lambda_1 \geq \ldots \geq \lambda_r > 0$ and $\lambda_{r+1} = \ldots = \lambda_n = 0$

Let $v_1, \ldots, v_r$ be the orthonormal vectors corresponding to $\lambda_1, \ldots, \lambda_r$

For $i = 1, \ldots, r$, define $\sigma_i = ||Av_i|| = \sqrt{\lambda_i}$
Note that $\sigma_1 \geq \ldots \sigma_r > 0$

For $i = 1, \ldots, r$, define $u_i = \frac{1}{\sigma_i} Av_i$

SVD for A

Vectors $u_1, \ldots, u_r$ are orthonormal and $Av_i = \sigma_i u_i$.

SVD of A

Summary

For a matrix A of dimension $m \times n$, where $m \geq n$, we have

1. A^TA is a symmetric positive semidefinite square matrix of dimension $n \times n$.
2. Rank of A is n: $\lambda_1 \geq \ldots \geq \lambda_n > 0$ are eigenvalues of A^TA and $v_1, \ldots, v_n$ the corresponding orthonormal eigenvectors. The vectors $u_i = Av_i/\sigma_i$, for $i = 1, \ldots, n$, are orthonormal, where $\sigma_i = \sqrt{\lambda_i}$.
3. Rank of A is $r < n$: $\lambda_1 \geq \ldots \geq \lambda_r > 0$ are non-zero eigenvalues of A^TA and $v_1, \ldots, v_r$ the corresponding orthonormal eigenvectors. The vectors $u_i = Av_i/\sigma_i$, for $i = 1, \ldots, r$, are orthonormal, where $\sigma_i = \sqrt{\lambda_i}$.
4. $AV = U\Sigma$, where V is $n \times r$ matrix consisting of orthonormal eigenvectors of A^TA corresponding to non-zero eigenvalues of A^TA, U is $m \times r$ matrix of orthonormal vectors given by $u_i = Av_i/\sigma_i$ for non-zero σ_i, and Σ is $r \times r$ diagonal matrix.
5. $AVV^T = A = U\Sigma V^T$

Matrices A^TA and AA^T

We have $A = U\Sigma V^T$.

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = (V\Sigma U^T)(U\Sigma V^T) = V\Sigma(U^TU)\Sigma V^T = V\Sigma^2V^T$$

Matrix A^TA

A^TA is square symmetric matrix and it is expressed in the diagonalized form $A^TA = V\Sigma^2V^T$. Thus, σ_i^2's are its eigenvalues and V is its eigenvectors matrix.

Similarly, consider AA^T and we obtain that
$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma V^TV\Sigma U^T = U\Sigma^2U^T$.

Matrix AA^T

AA^T is square symmetric matrix and it is expressed in the diagonalized form $AA^T = U\Sigma^2U^T$. Thus U is the eigenvector matrix for the symmetric matrix AA^T with the same eigenvalues as A^TA.

Singular Value Decomposition

- Let A be a $m \times n$ matrix of real numbers of rank r

- $A_{m\times n} = U_{m\times r}\Sigma_{r\times r}V^T_{r\times n}$, where

U is a orthonormal $m \times r$ matrix
V is a orthonormal $n \times r$ matrix
Σ is an $r \times r$ diagonal matrix and its (i, i)-th entry is σ_i for $i = 1, \ldots, r$

- Note that $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$ and $\sigma_i = \sqrt{\lambda_i}$ where λ_i are the eigenvalues of $A^T A$

- The set of orthonormal vectors $v_1, \ldots, v_r$ and $u_1, \ldots, u_r$ are eigenvectors of $A^T A$ and AA^T, respectively. The vectors v's and u's satisfy the equation $Av_i = \sigma_i u_i$, for $i = 1, \ldots, r$

- Alternatively, we can express A as the sum of the product of rank 1 matrices

$$A = \Sigma_{i=1}^r \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$$

Applications

An Application

Let $A_{m \times n}$ be the Utility Matrix, where $m = 10^8$ users and $n = 10^5$ items.

SVD of $A = U\Sigma V^T$

Let r of $\sigma_i' s$ are > 0
Let $\sigma_1 \geq \ldots \geq \sigma_r > 0$

A can be expressed as $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$

Total space required to store A is $r(m + n + 1)$. If rank of A is small, it is better to store $u_1, \ldots, u_r, v_1, \ldots, v_r, \sigma_1, \ldots, \sigma_r$, rather than whole of A.

Low Rank Approximation

Energy of A is given by $\mathcal{E} = \sum_{i=1}^{r} \sigma_i^2$

Define $\mathcal{E}' = 0.99\mathcal{E}$, and let $j \leq r$ be the maximum index such that $\sum_{1=1}^{j} \sigma_i^2 \leq \mathcal{E}'$

Approximate A by $\sum_{i=1}^{j} \sigma_i u_i v_i^T$

How many cells we need to store in this representation?

1. First j columns of U,
2. j diagonal entries of Σ, and
3. j rows of V^T.

Total Space = $j^2 + j(m+n)$ cells

Low Rank Approximation (contd.)

For our example, dimension of $A_{m \times n}$ are $m = 10^8$ users and $n = 10^5$ items.

If $j = 20$, then we need to store
$j^2 + j(m+n) = 20^2 + 20 \times (10^8 + 10^5) \approx 5,005,000$ cells

This number is only $.02\%$ of 10^{13}

Low Rank Approximations

Let SVD of A be

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{30} & -1/\sqrt{5} \\ 1/\sqrt{30} & 2/\sqrt{5} \\ 5/\sqrt{30} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

In terms of Rank 1 Components:

$$A = \sqrt{6} \begin{bmatrix} 2/\sqrt{30} \\ 1/\sqrt{30} \\ 5/\sqrt{30} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}^T + \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}^T$$

Energy of A: $\mathcal{E}(A) = \sqrt{6}^2 + 1^2 = 7$

Possible $\frac{6}{7}$-Energy approximation of A is given by

$$A \approx \sqrt{6} \begin{bmatrix} 2/\sqrt{30} \\ 1/\sqrt{30} \\ 5/\sqrt{30} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}^T$$

Interpreting $U, \Sigma,$ and V

Utility Matrix M as SVD $M = U\Sigma V^T$

$$
\text{M=}\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} =
$$

$$
\begin{bmatrix} .13 & -.02 & .01 \\ .41 & -.07 & .03 \\ .55 & -.1 & .04 \\ .68 & -.11 & .05 \\ .15 & .59 & -.65 \\ .07 & .73 & .67 \\ .07 & .29 & -.32 \end{bmatrix} \begin{bmatrix} 12.5 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.35 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \\ .40 & -.8 & .40 & .09 & .09 \end{bmatrix}
$$

1. 3 concepts $(= rank)$
2. U maps users to concepts
3. V maps items to concepts
4. Σ gives strength of each concept

Rank-2 Approximation

$$\begin{bmatrix} .13 & -.02 \\ .41 & -.07 \\ .55 & -.1 \\ .68 & -.11 \\ .15 & .59 \\ .07 & .73 \\ .07 & .29 \end{bmatrix} \begin{bmatrix} 12.5 & 0 \\ 0 & 9.5 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \end{bmatrix}$$

% Loss in Energy $= \frac{1.35^2}{12.5^2+9.5^2+1.35^2} < 1\%$

Mapping Users to Concept Space

Consider the utility matrix M and its SVD.

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} .13 & -.02 \\ .41 & -.07 \\ .55 & -.1 \\ .68 & -.11 \\ .15 & .59 \\ .07 & .73 \\ .07 & .29 \end{bmatrix} \begin{bmatrix} 12.5 & 0 \\ 0 & 9.5 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \end{bmatrix}$$

MV gives mapping of each user in concept space:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} .56 & -.12 \\ .59 & .02 \\ .56 & -.12 \\ .09 & .69 \\ .09 & .69 \end{bmatrix} = \begin{bmatrix} 1.71 & -.22 \\ 5.13 & -.66 \\ 6.84 & -.88 \\ 8.55 & -1.1 \\ 1.9 & 5.56 \\ .9 & 6.9 \\ .96 & 2.78 \end{bmatrix}$$

Mapping Users to Items

Suppose we want to recommend items to a new user q with the following row in the utility matrix $\begin{bmatrix} 4 & 0 & 0 & 0 & 0 \end{bmatrix}$

1. Map q to concept space:

$$qV = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} .56 & -.12 \\ .59 & .02 \\ .56 & -.12 \\ .09 & .69 \\ .09 & .69 \end{bmatrix} = \begin{bmatrix} 1.68 & -.36 \end{bmatrix}$$

2. Map the vector qV to the Items space by multiplying by V^T as vector V captures the connection between items and concepts.

$$\begin{bmatrix} 1.68 & -.36 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \end{bmatrix} =$$
$$\begin{bmatrix} .98 & .98 & .98 & -.1 & -.1 \end{bmatrix}$$

Mapping Users to Items (Contd.)

Suppose we want to recommend items to user q' with the following row in the utility matrix $\begin{bmatrix} 0 & 0 & 0 & 4 & 0 \end{bmatrix}$

1. $q'V = \begin{bmatrix} 0 & 0 & 0 & 4 & 0 \end{bmatrix} \begin{bmatrix} .56 & -.12 \\ .59 & .02 \\ .56 & -.12 \\ .09 & .69 \\ .09 & .69 \end{bmatrix} = \begin{bmatrix} .36 & 2.76 \end{bmatrix}$
2. Map $q'V$ to the Items space by multiplying by V^T
   $\begin{bmatrix} .36 & 2.76 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \end{bmatrix} =$
   $\begin{bmatrix} -.12 & .26 & -.12 & 1.93 & 1.93 \end{bmatrix}$

Mapping Users to Items (Contd.)

Suppose we want to recommend items to user q'' with the following row in the utility matrix $\begin{bmatrix} 0 & 0 & 4 & 4 & 0 \end{bmatrix}$

1. $q''V = \begin{bmatrix} 0 & 0 & 4 & 4 & 0 \end{bmatrix} \begin{bmatrix} .56 & -.12 \\ .59 & .02 \\ .56 & -.12 \\ .09 & .69 \\ .09 & .69 \end{bmatrix} = \begin{bmatrix} 2.6 & 2.28 \end{bmatrix}$
2. Map $q''V$ to the Items space by multiplying by V^T

   $\begin{bmatrix} 2.6 & 2.28 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \end{bmatrix} =$

   $\begin{bmatrix} 1.18 & 1.57 & 1.18 & 1.8 & 1.8 \end{bmatrix}$

CUR

Interpreting $U, \Sigma,$ and V

Utility Matrix M as SVD $M = U\Sigma V^T$

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} .13 & -.02 & .01 \\ .41 & -.07 & .03 \\ .55 & -.1 & .04 \\ .68 & -.11 & .05 \\ .15 & .59 & -.65 \\ .07 & .73 & .67 \\ .07 & .29 & -.32 \end{bmatrix} \begin{bmatrix} 12.5 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.35 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \\ .40 & -.8 & .40 & .09 & .09 \end{bmatrix}$$

Issues:

1. Utility matrix M is sparse, but U and V are dense
2. Total size $= r(n+m) + r^2$
3. Interpretation of entries of U and V is unclear

A Possiblity

Can we express $M \approx \mathcal{C}U\mathcal{R}$, where

1. $\mathcal{C}$ consists of some columns of M
2. $\mathcal{R}$ consists of some rows of M
3. U is not that big
4. Square of Frobenius Norm $= \sum_{ij}(M_{ij} - (\mathcal{C}U\mathcal{R})_{ij})^2$ is small

$\mathcal{C}$U$\mathcal{R}$ Method

Let M be $m \times n$ and let $\Delta = \sum_{ij} M[i,j]^2$

1. For each column j, compute $p_j = \frac{1}{\Delta} \sum_{i=1}^{m} M[i,j]^2$
2. Pick α columns of M based on their probabilities (with replacement). Let $\mathcal{C}$ be the multi-set of picked columns.
3. For each element of selected columns $j \in \mathcal{C}$, scale its value to $\frac{M[*,j]}{\sqrt{\alpha p_j}}$
4. Repeat above steps for all the rows and let $\mathcal{R}$ be the multi-set of α picked and scaled rows.
5. Let W be the $\alpha \times \alpha$ matrix whose entries are from M that are common to $\mathcal{C}$ and $\mathcal{R}$
6. Construct SVD of $W = X \Sigma Y^T$. Construct Σ^+, where each non-zero element x of Σ is replaced by $1/x$
7. Compute $U = Y(\Sigma^+)^2 X^T$
8. Report $\mathcal{C}U\mathcal{R}$ as approximation of M

An Example

$$M = \begin{bmatrix} 4 & 1 & 1 & 0 \\ 4 & 0 & 0 & 1 \\ 0 & 0 & 5 & 5 \\ 0 & 1 & 5 & 5 \\ 0 & 1 & 5 & 3 \end{bmatrix} \qquad \sum_{ij} {M_{ij}}^2 = 171$$

Column	C_1	C_2	C_3	C_4	
$\sum_i M_{ij}$	32	3	76	60	
Row	R_1	R_2	R_3	R_4	R_5
$\sum_j M_{ij}$	18	17	50	51	35

For Rank 1 Approximation: Select C_3 and R_4 and scale them to obtain:

$$C_3 = \frac{1}{\sqrt{76/171}} \begin{bmatrix} 1 & 0 & 5 & 5 & 5 \end{bmatrix}^T = \begin{bmatrix} 1.5 & 0 & 7.6 & 7.6 & 7.6 \end{bmatrix}^T$$

$$R_4 = \frac{1}{\sqrt{51/171}} \begin{bmatrix} 0 & 1 & 5 & 5 \end{bmatrix} = \begin{bmatrix} 0 & 1.85 & 9.3 & 9.3 \end{bmatrix}$$

$W = 5$ and SVD $W = [1][5][1]$ and $U = [1][1/25][1]$

An Example (contd.)

We have $M = \mathcal{C}U\mathcal{R}$ as

$$M \approx \begin{bmatrix} 1.5 \\ 0 \\ 7.6 \\ 7.6 \\ 7.6 \end{bmatrix} \begin{bmatrix} \frac{1}{25} \end{bmatrix} \begin{bmatrix} 0 & 1.85 & 9.3 & 9.3 \end{bmatrix} = \begin{bmatrix} 0 & .11 & .59 & .59 \\ 0 & 0 & 0 & 0 \\ 0 & .56 & 2.8 & 2.82.8 \\ 0 & .56 & 2.8 & 2.82.8 \\ 0 & .56 & 2.8 & 2.82.8 \end{bmatrix}$$

Try Rank 2 approximation: Possibly select Columns C_3, C_4 and Rows R_3, R_4 and compute scaled columns, rows, matrices W, U and $\mathcal{C}U\mathcal{R}$

Remarks

1. In case a row/column is picked $\beta > 1$ times, we take only one of its copy in $\mathcal{R}/\mathcal{C}$ and scale the corresponding entries by a factor of $\sqrt{\beta}$
2. $\implies W$ may not be square, but we know how to compute SVDs for rectangular matrices.
3. Columns in $\mathcal{C}$ and rows in $\mathcal{R}$ are from M
4. In $\mathcal{C}U\mathcal{R}$ decomposition, U (of dimension at most $\alpha \times \alpha$) may be dense.
5. Total Space $= \alpha(n+m) + \alpha^2$ (likely to be much less due to the sparsity of M)

Performance Guarantee

Quality Estimate

Let M_k be the best rank k-approximation of M.
Choose $\alpha = \frac{k \log k}{\epsilon^2}$.
The resulting $\mathcal{C}U\mathcal{R}$ decomposition satisfies the following:
Frobenius Norm of M and $\mathcal{C}U\mathcal{R}$ is at most $(2+\epsilon)$ times the Frobenius Norm of M and M_k, i.e. $||M - \mathcal{C}U\mathcal{R}||_F \leq (2+\epsilon)||M - M_k||_F$

Remarks:

1. There are recent works that show that $\alpha = k/\epsilon$ suffices
2. Approximation is by a factor of $1+\epsilon$
3. Running time is faster than that of computing SVDs
4. Randomized Linear Algebra - a new field in TCS

References

1. Gilbert Strang, Introduction to Linear Algebra, Wellesley-Cambridge Press.
2. G. H. Golub abd W. Kahan, Calculating the singular values and pseudo-inverse of a matriix, SIAM Journal Series B2:2:205-224, 1965.
3. P. Drineas, R. Kanan and M.W. Mahoney, Fast Monte-Carlo algorithms for matrics III: Computing a compressed approximate matrix decomposition, SIAM Jl. Computing 36:1: 184-206, 2006.
4. J. Sun, Y. Xie, H. Zhang, and C. Faloutsos, Less is more: compact matrix decomposition for large sparse graphs, SIAM International Conference on Data Mining, 2007.