Singular Value Decomposition

An application to Big Data

Davide Sferrazza

January 10, 2023

Università degli Studi di Palermo

Outline

1. Singular Value Decomposition
   - What is it?
   - How can singular values be computed?
2. Finding eigenvalues and eigenvectors
   - QR Method
   - Improving QR Method using Givens matrices
3. References

Singular Value Decomposition

Definition of SVD

Theorem

Given a matrix $A \in \mathbb{R}^{m \times n}$, it can always be found a decomposition such that

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

where $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ and $\Sigma \in \mathbb{R}^{m \times n}$.

U and V are two orthogonal matrices and Σ is a diagonal matrix, namely:

$$(\Sigma)_{ij} = \begin{cases} 0, & i \neq j \\ \sigma_i, & i = j \end{cases}$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$, $p = \min\{m, n\}$.

Definition of SVD

The non-zero entries of Σ, denoted by σ_i, are called *singular values*.

They are arrenged in a nonincreasing order by convention.

The column vectors $\boldsymbol{u_i}$ of U are called *left singular vectors* and those $\boldsymbol{v_i}$ of V are called *right singular vectors*.

Since in general $m \neq n$, we have:

$$A = \sum_{i=1}^{p} \boldsymbol{u_i} \sigma_i \boldsymbol{v_i}^T$$

Definition of SVD

Theorem

If for some r such that $1 \leq r < p$ we have

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0$$

then

- $rank(A) = r$
- $A = \sum_{i=1}^{r} \boldsymbol{u_i} \sigma_i \boldsymbol{v_i}^T$

This means that all other $p - r$ dimensions of matrix A are linear combinations of the first r.

Definition of SVD

Lower rank approximation

Let $A \in \mathbb{R}^{m \times n}$ be a matrix whose rank is $rank(A) = r$.
If for a fixed integer value $k < r$ we define

$$A_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u_i} \boldsymbol{v_i}^T \tag{1}$$

and

$$\mathcal{B} = \left\{ B \in \mathbb{R}^{m \times n} : rank(B) = k \right\}$$

then

$$\min_{B \in \mathcal{B}} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

This result tell us that A_k represents the best approximation (considering the *spectral norm*) of rank k of matrix A.

Singular values computation

To compute the singular values, consider the transponse of A given its decomposition:

$$A^T = (U \Sigma V^T)^T = V \Sigma^T U^T$$

The symmetric matrix $A^T A$ is equal to:

$$A^T A = (V \Sigma^T U^T)(U \Sigma V^T) = V \Sigma^T \Sigma V^T$$

Furthermore, this equation can be written as:

$$A^T A V = V \Sigma^T \Sigma$$

This means that the diagonal entries of the square matrix $\Sigma^T \Sigma$, which are the square of the singular values, are the eigenvalues of matrix $A^T A$ and V is the matrix of eigenvectors.

Singular values computation

Similarly, consider the product of AA^T. It is equal to:

$$AA^T = (U\Sigma V^T)(V\Sigma^T U^T) = U\Sigma\Sigma^T U^T$$

Which means that:

$$AA^T U = U\Sigma\Sigma^T$$

Hence U is the matrix of eigenvectors of AA^T.

Since $rank(A) = r$, only the first r eigenvalues of AA^T and $A^T A$ are non-zero.

Finding eigenvalues and eigenvectors

QR Method

A possible method to find eigenvalues and eigenvectors of a matrix is based on QR decompositions and this theorem:

Theorem

Suppose $A \in \mathbb{R}^{n \times n}$ *is a matrix having eigenvalues* $\lambda_1, \lambda_2, \ldots \lambda_n$ *satisfying*

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| \tag{2}$$

then the following sequence for $A_1 = A$ *and* $k = 1, 2, \ldots$

$$\begin{cases} A_k = Q_k R_k \\ A_{k+1} = R_k Q_k \end{cases} \tag{3}$$

converges to an upper triangular matrix where $(A_k)_{ii} = \lambda_i,\ i = 1, 2, \ldots, n.$
In case (2) is not satisfied, this sequence converges to a triangular matrix with square blocks of order at most 2 along the diagonal.
If A *is symmetric, then the sequence converges to a diagonal matrix.*

QR Method

So a basic implementation would be like this:

```
while err < toll
    [Q, R] = qr(A);
    A = R * Q;

    err = max( max( tril(A, -1) ) );
end
```

This method could be speed up using a technique called *shifing*:

```
n = length( A );
while err < toll
    % A(n, n) is an usual choice, it could be any real number
    T = A(n, n) * eye(n);
    [Q, R] = qr( A - T );
    A = R * Q + T;

    err = max( max( tril(A, -1) ) );
end
```

QR Method

In our case, this method must be applied to AA^T and A^TA, so if (3) converges then we have a diagonal matrix.
If we consider the diagonalization of B, where $B = AA^T$ or $B = A^TA$:

$$B = P\Lambda P^{-1} = P\Lambda P^T$$

As B can be factorized using (3), the matrix containing the eigenvectors must be equal to:

$$P = \prod_i Q_i = Q_1 Q_2 Q_3 \cdots$$

Hence, for every iteration $B_k = Q_k R_k$ and $B_{k+1} = R_k Q_k$, requiring each step to have a computational cost equal to $O(\frac{2n^3}{3})$.

Hessember Reduction

Prova i

```
function [c, s] = givens(A, i, j)
%GIVENS Function to compute cos and sin of Gij
%   Given a matrix A, this function returns a vector containing the values
%   of givens matrix Gij such that Gij * A is equal to A except for the
%   element (i, j), which will be zero.
     c = abs( A(i, i) ) / sqrt( A(i, i)^2 + A(j, i)^2 );
     s = - sign( A(j, i) / A(i, i) ) * abs( A(j, i) ) / sqrt( A(i, i)^2 + A
     (j, i)^2 );
end
```

References

References

- Aggarwal, Charu C et al. *Data mining: the textbook*. Vol. 1. Springer, 2015.
- Dua, Dheeru and Casey Graff. *UCI Machine Learning Repository*. 2017. URL: http://archive.ics.uci.edu/ml.
- Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. *Mining of massive data sets*. Cambridge university press, 2020.
- Monegato, Giovanni. *Metodi e algoritmi per il calcolo numerico*. Clut, 2008.
- Van Loan, Charles F and G Golub. "Matrix computations (Johns Hopkins studies in mathematical sciences)". In: *Matrix Computations* (1996).

References ii

Yılmaz, Nevriye and Boran Sekeroglu. “Student Performance Classification Using Artificial Intelligence Techniques”. In: *10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019*. Ed. by Rafik A. Aliev et al. Cham: Springer International Publishing, 2020, pp. 596–603. ISBN: 978-3-030-35249-3.