Singular Value Decomposition

Tuan Nguyen

Ngày 9 tháng 1 năm 2021

Overview

A diagonal matrix

A diagonal matrix is a matrix that has value zero on all off-diagonal elements

$$D = \begin{bmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_n \end{bmatrix}$$

They allow fast computation of determinants, powers, and inverses.

$$D^{-1} = \begin{bmatrix} \frac{1}{c_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{c_n} \end{bmatrix}, D^k = \begin{bmatrix} c_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_n^k \end{bmatrix}$$

A diagonal matrix (cont.)

A matrix $A \in \mathrm{R}^{n \times n}$ is **diagonalizable** if there exists an invertible matrix $P \in \mathrm{R}^{n \times n}$ such that $D = P^{-1}AP$

We define $P := [p_1, ..., p_n]$ and let $D \in \mathrm{R}^n$ be a diagonal matrix with diagonal entries $\lambda_1, ..., \lambda_n$

$$AP = A[p_1, ..., p_n] = [Ap_1, ..., Ap_n]$$

$$PD = [p_1, ..., p_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = [\lambda_1 p_1, ..., \lambda_n p_n]$$

$$PD = AP \Leftrightarrow [Ap_1, ..., Ap_n] = [\lambda_1 p_1, ..., \lambda_n p_n]$$

$$\Leftrightarrow Ap_1 = \lambda_1 p_1, ..., Ap_n = \lambda_n p_n$$

Therefore, the columns of P must be eigenvectors of A and the diagonal of D are the eigenvalues of A.

Eigendecomposition

A square matrix $A \in \mathrm{R}^{n \times n}$ can be factored into

$$A = PDP^{-1}$$

where $P \in \mathrm{R}^{n \times n}$ and D is a diagonal matrix whose diagonal entries are the eigenvalues of A, if and only if the eigenvectors of A form a basis of R^n

Remark: A symmetric matrix $S \in \mathrm{R}^{n \times n}$ can always be diagonalized.

Find eigendecomposition of

$$A = \frac{1}{2}\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$$

Eigendecomposition (cont.)

:

- Diagonal matrices D can efficiently be raised to a power. Therefore, we can find a matrix power for a matrix $A \in \mathrm{R}^{n \times n}$ via the eigenvalue decomposition (if it exists)

$$A^k = (PDP^{-1})^k = PD^k P^{-1}$$

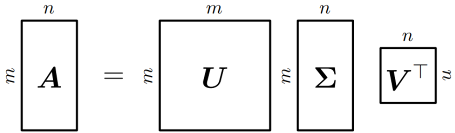
- Assume that the eigendecomposition $A = PDP^{-1}$ exists. Then

$$det(A) = det(PDP^{-1}) = det(P)det(D)det(P^{-1}) = det(D) = \prod_{i=1}^{n} d_{ii}$$

However, the eigenvalue decomposition requires square matrices. It would be useful to perform a decomposition on general matrices => Singular Value Decomposition.

Singular Value Decomposition

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



with an orthogonal matrix $U \in \mathrm{R}^{m \times m}$ with column vectors u_i, i = 1,...,m, and an orthogonal matrix $V \in \mathrm{R}^{n \times n}$ with column vectors v_j, j = 1,...,n. Moreover, Σ is an m $\times$ n matrix with $\Sigma_{ii} = \sigma_i > 0$ and $\Sigma_{ij} = 0$ if $i \neq j$.

The diagonal entries θ_i, i = 1,...,r, of Σ are called the singular values, u_i are called the left-singular vectors, and v_j are called the right-singular vectors. By convention, the singular values are ordered, i.e., $\theta_1 > \theta_2 > ...$

Singular Value Decomposition (cont.)

Remark: Θ has the same size with A.

- If m > n

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

- If n > m

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \sigma_m & 0 & \cdots & 0 \end{bmatrix}$$

The SVD exists for any matrix $A \in R^{m \times n}$

$$A = U\Sigma V^T \Rightarrow A^T A = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T =$$
$$V\Sigma^T\Sigma V^T = V \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix} V^T$$

We know that $A^T A$ is symmetric, positive semidefinte, thus we can diagonalize $A^T A$

$$A^T A = PDP^T = P \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} P^T$$
$$\Rightarrow V = P, \sigma_i = \sqrt{\lambda_i}, i = 1, .., n$$

Therefore, the eigenvectors of $A^T A$ that compose P are the right-singular vectors V of A. The eigenvalues of $A^T A$ are the squared singular values of Σ.

$$A = U\Sigma V^T \Rightarrow AA^T = U\Sigma V^T(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T = U \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix} U^T$$

AA^T is symmetric and can be diagonalized, then the orthonormal eigenvectors of AA^T are the left-singular vectors U and form an orthonormal basis in the codomain of the SVD.

Finally, AA^T and A^TA have the same eigenvalues.

Construction of SVD (cont.)

How to construct SVD for a matrix A

$A = U\Sigma V^T \Leftrightarrow AV = U\Sigma \Leftrightarrow Av_i = u_i\sigma_i \Leftrightarrow u_i = \dfrac{1}{\sigma_i}Av_i$ Steps:

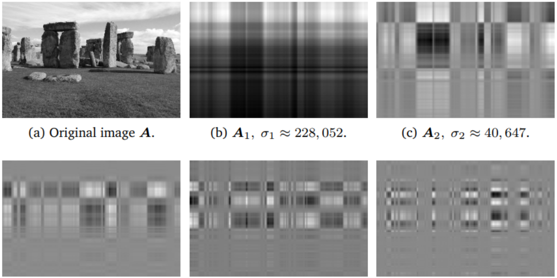
- Find the normalize eigenvalue and eigenvalue of $A^T A$
- Construct V
- Find the U $u_i = \dfrac{1}{\sigma_i}Av_i$

Find the SVD of A, where:

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$$

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T, \mathrm{r} = \mathrm{rank(A)}$$

We can construct rank-1 matrix $A_i \in \mathrm{R}^{m\times n}$ as $A_i = u_i v_i^T$



(a) Original image $\boldsymbol{A}$.
(b) $\boldsymbol{A}_1,\ \sigma_1 \approx 228,052$.
(c) $\boldsymbol{A}_2,\ \sigma_2 \approx 40,647$.
(d) $\boldsymbol{A}_3,\ \sigma_3 \approx 26,125$.
(e) $\boldsymbol{A}_4,\ \sigma_4 \approx 20,232$.
(f) $\boldsymbol{A}_5,\ \sigma_5 \approx 15,436$.

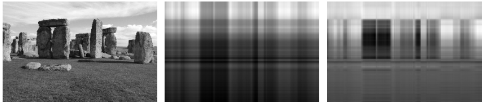
Hình 1: Image construction by rank-1 matrix

Matrix Approximation (cont.)

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \lambda_i A_i$$

We can construct rank-k approximation of A

$$\hat{A}(k) = \sum_{i=1}^{k} \sigma_i u_i v_i^T = \sum_{i=1}^{k} \lambda_i A_i$$



(a) Original image $\boldsymbol{A}$. (b) Rank-1 approximation $\widehat{\boldsymbol{A}}(1)$. (c) Rank-2 approximation $\widehat{\boldsymbol{A}}(2)$.

(d) Rank-3 approximation $\widehat{\boldsymbol{A}}(3)$. (e) Rank-4 approximation $\widehat{\boldsymbol{A}}(4)$. (f) Rank-5 approximation $\widehat{\boldsymbol{A}}(5)$.

Matrix Approximation (cont.)

We can interpret the rank-k approximation obtained with the SVD as a

- projection of the full-rank matrix A onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between A and any rank-k approximation.
- rank-k matrix as a form of lossy compression. Therefore, the low-rank approximation of a matrix appears in many machine learning applications, e.g., image processing, noise filtering, and regularization of ill-posed problems. Furthermore, it plays a key role in dimensionality reduction and principal component analysis