Singular Value Decomposition
04/28/2025

Singular value decomposition is a linear algebra technique to decompose a real number matrix, not necessarily a square matrix, into a diagonal representation. The diagonal positive entries of the latter matrix are the "singular values" ranked by their magnitudes. This representation is intended to captures the "principal components" of the original matrix therefore has found broad applications in data compression, image processing, and algorithms that require a "concentrated" inputs without losing much information carried by the original data.

Smith Canonical Form

Let A be an m x n matrix of rank r, then there exists invertible matrices U and V of size m x m and n x n, respectively, such that

$$UAV = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$$

The matrix on the right is the Smith canonical form of matrix A. Further, if matrix R is the reduced row-echelon form of A, then U is computed by $[A\ I_m] \Rightarrow [R\ U]$ and similarly V is computed by

$$\left[R^T\ I_n\right] \Rightarrow \left[\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times m} V^T\right]$$

Matrix R is the closest matrix to the identity matrix *by row operations* of A, whereas the Smith canonical form is the closest matrix to the identity matrix by *both row and column operations* of A. In fact, for a square, invertibale matrix A, *i.e.,* m = n = r, R and V become the identity matrix, and U the inverse of A.

Diagonalization

Let A be an n x n matrix, then A is diagonalizable if and only if it has eigenvectors, $\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n$, such that P = $[\mathbf{p}_1\ \mathbf{p}_2\ \cdots\ \mathbf{p}_n]$ is invertible. Further, in this case, $P^{-1}AP = diag(\lambda_{1,2}, \ldots, \lambda_n)$ where λ_i is the eigenvalue of A corresponding to $\mathbf{p}_i$.

Gram-Schmidt Orthogonalization

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a basis of a subspace S of $\mathbb{R}^n$, an orthonormal basis of S, $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$, can be constructed successively as follows:

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{x}_1 \\
\mathbf{y}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{y}_1}{\|\mathbf{y}_1\|^2}\mathbf{y}_1 \\
\mathbf{y}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{y}_1}{\|\mathbf{y}_1\|^2}\mathbf{y}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{y}_2}{\|\mathbf{y}_2\|^2}\mathbf{y}_2 \\
&\vdots
\end{aligned}$$

If a square matrix is symmetric, the column vectors form an orthogonal basis for the column subspace. Gram-Schmidt orthogonalization is not required. In other cases, where some eigenvalues are of multiplicity, the Gram-Schmidt procedure needs to be performed on these eigenvectors to make them orthogonal.

Singular Value Decomposition

Let A be a real m x n matrix, then A^TA and AA^T have the same set of positive, real eigenvalues, λ_i, corresponding to the eigenvectors, $\mathbf{q}_i$, that form the column vectors of n x n matrix Q.

Further, the real numbers $\sigma_i = \sqrt{\lambda_i}$ are called the *singular values* of matrix A.

Let r be the rank of matrix A, the *singular matrix* of A is defined as below:

$$\Sigma_A = \begin{bmatrix} D_A & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$$

where $D_A = diag\left(\sigma_1, \sigma_2, \ldots, \sigma_r\right)$. Matrix A can be factorized as below:

$$A = P\Sigma Q^T$$

This is called *singular value decomposition* of matrix A, and P is an m x m orthogonal matrix with the first r column vectors constructed as below:

$$\mathbf{p}_i = \frac{1}{\|A\mathbf{q}_i\|} A\mathbf{q}_i$$

The subsequent vectors are constructed following the Gram-Schmidt procedure to form the orthonormal basis.

Central Theorem of Linear Algebra

Any matrix A has four subspaces: the row space, column space, null space of A and finally, the null space of A^T. One can construct basis for each of these subspaces. However, we would want the bases to be orthonormal and the matrix with respect to these bases diagonal. Let A be an m x n matrix with rank r. Since A^TA is symmetric and positive semidefinite, it has nonnegative eigenvalues (σ_i^2) and orthonormal eigenvectors (v_i) thus form a basis for the row space of A^TA.

$A^TAv_i = \sigma_i^2 v_i$

$v_i^TA^TAv_i = \sigma_i^2 v_i^T v_i$, therefore $\|Av_i\| = \sigma_i$

$AA^TAv_i = \sigma_i^2 Av_i$

Av_i is an eigenvector of A^TA and $u_i = Av_i/\sigma_i$ is a unit eigenvector of A^TA.

Further, if $v_1, v_2, \ldots, v_r$ form the basis of the row space and $u_1, u_2, \ldots, u_r$ form the basis of the column space of A, then

$$Av_i = \sigma_i u_i$$

This means that $AV = \Sigma U$. Then right multiplying by V^T, we have

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

When A itself is symmetric, the eigenvectors u_i diagonalize A. The factorization of $A = U\Sigma V^T$ together with $A = LU$ and $A = QR$ form the central theorem in linear algebra.

Reduced SVD and Pseudo-Inverse

When Σ contains zeros, *i.e.,* r < m, n, a more compact decomposition of A is possible. U and V can be partitioned as follows:

$$U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$$
$$V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

The partitioned matrix multiplication becomes:

$$A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T$$

This is called reduced singular value decomposition of A. Now, since the diagonal entries in D are all positive and non-zero, D is invertible.

$$A^+ = V_r D^{-1} U_r^T$$

This is called Moore-Penrose inverse of A, or pseudo-inverse.

Least Square Solution of Linear Systems

Given a linear systems equation $A\mathbf{x} = \mathbf{b}$, then the least square solution is

$$\hat{\mathbf{x}} = A^+\mathbf{b}$$

We can use the definition of A^+ and SVD of A to show:

$$\begin{aligned} A\hat{\mathbf{x}} &= \left(U_r D V_r^T\right)\left(V_r D^{-1} U_r^T \mathbf{b}\right) \\ &= U_r D D^{-1} U_r^T \mathbf{b} \\ &= U_r U_r^T \mathbf{b} \end{aligned}$$

Therefore, $\hat{\mathbf{x}}$ is a least square solution of $A\mathbf{x} = \mathbf{b}$.

Image Compression

Let A_k be a k-rank approximation of an image matrix A, such that

$$A_k = U_k \Sigma_k V_k^T$$

where U_k is the k-rank approximation of U that only has the first k columns of U, Σ_k only the first k singular values of Σ, and V_k only the first k rows of V. The Euclidian distance between A and A_k is:

$$\|A - A_k\|^2 = \sum_{i=1}^{r} (\sigma_i - \sigma_{ki})^2$$

Therefore, if only the smallest entries in Σ are set to zero, the loss caused by using the approximation image matrix is minimized. On the other hand, the savings in data storage of A_k can be substantial in comparison with the original image matrix.

Online Recommendation System

Let A be an m x n "ranking matrix" whose rows are ratings of each user for products that form the columns of the matrix. The SVD of A yields matrices U, Σ, V, where matrix U transforms each user to the product "categories", Σ the "strength" of each category as perceived by the users, and V transforms each category to a group of products. The recommendation system then computes the category based on a particular user's ratings followed by recommending other products of the same category to the user.

Facial Recongnition

Let A and B be image matrices of dimensions m x n. The inner product of the two matrices is

$$\langle A, B \rangle = tr\left(B^T A\right)$$

The similarity of these matrices is defined as the cosine angle between the two:

$$\alpha(A, B) = \cos(\theta_{A,B}) = \frac{\langle A, B \rangle}{\|A\| \, \|B\|}$$

Upon performing the SVD of A and B, the overall similarity of the two images is determined by the sum of the similarities of the matrices U and V:

$$\beta_{A,B} = \alpha_U(A, B) + \alpha_V(A, B)$$

The facial recognition criterion is then the "correlation" of the two images:

$$\rho_{A,B} = \frac{\beta_{A,B}}{d_{A,B}}$$

where $d_{A,B}$ is the Euclidian distance of their singular values:

$$d_{A,B} = \|A - B\|$$

The likeliness of two facial images, *i.e.*, recognition, is then determined by their correlation coefficient.