4.5 Singular Value Decomposition(특이값분해)

SVD: a central matrix decomposition method in linear algebra

Applied to all matrices, and always exists

고유값 분해처럼 행렬을 대각화하는 한 방법

$A = U\Sigma V^T$ 로 분해가 된다.

이 때, U 의 열벡터는 AA^T 의 고유 벡터, V 의 열벡터는 A^TA 로 구성된다.

또한 $\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \\ 0 & \cdots & 0 \end{bmatrix}$ $(m > n)$, $\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & & & 0 \\ & \ddots & & \vdots \\ & & \sqrt{\lambda_n} & 0 \end{bmatrix}$ $(n < m)$ 로 정의된다.

이 때, λ_k 는 k 번째 AA^T 및 A^TA 의 고유값이다.

4.5.1 Geometric Intuitions for the SVD

SVD performs a basis change via V transpose followed by a scaling and augmentation/reduction in dimensionality via singular value matrix

Then, it performs a second basis change vis U

Eigendecomposition operates within same vector space, while SVD expresses a change of basis in both the domain and codomain

Two different bases are simultaneously linked by the singular value matrix

4.5.2 Construction of the SBD

-Eigendecomposition of an SPD matrix

S = S transpose = P*D*P transpose

Corresponding SVD

S = U ΣV transpose

If we set

U = P = V, D = Σ

We see that the SVD of SPD matrices is their eigendecomposition.

Thus, the eigenvectors of Atranspose*A which we know are the right-singular vecotrs vi, and their normalized images under A, the left-singular vectors ui, form two self-consistent ONBs that are connected through the singular value matrix Σ.

Let us rearrange to obtain the singular value equation

4.5.3 Eigenvalue Decomposition vs SVD

Key difference: In the SVD, domain and codomain can be vector spaces of different dimensions

For convenience in notation and abstraction, we use an SBD notation where SVD is described as having two square left- and right-singular vector matrices, but a non-square singular value matrix. Our definition for the SVD is sometimes called the full SVD.

4.6 Matrix Approximation

-Investigate how the SBD allows us to represent a matrix A as a sum of simpler low-rank matrices Ai

-*Eckart-Young Theorem* states explicitly how much error we introduce by approzimating A using a rank-k approximation.

-As a projection of the full-rank matrix A onto a lower-dimensional space of rank-at-most-k matrices.

-SVD minimizes the error between A and any rank-k approximation.

+) Implies that we can use SVD to reduce a rank-r matrix A to a rank-k matrix Â in a principled, optimal manner.

Definition 4.23 (Spectral Norm of a Matrix)**.** For $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, the *spectral norm* of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is defined as

$$\|\boldsymbol{A}\|_2 := \max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}\,. \tag{4.93}$$

We introduce the notation of a subscript in the matrix norm (left-hand side), similar to the Euclidean norm for vectors (right-hand side), which has subscript 2. The spectral norm (4.93) determines how long any vector $\boldsymbol{x}$ can at most become when multiplied by $\boldsymbol{A}$.

Theorem 4.24. *The spectral norm of $\boldsymbol{A}$ is its largest singular value σ_1.*

We leave the proof of this theorem as an exercise.

Theorem 4.25 (Eckart-Young Theorem (Eckart and Young, 1936))**.** *Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ of rank r and let $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ be a matrix of rank k. For any $k \leqslant r$ with $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ it holds that*

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(\boldsymbol{B})=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2\,, \tag{4.94}$$

$$\left\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\right\|_2 = \sigma_{k+1}\,. \tag{4.95}$$

4.7 Matrix Phylogeny (행렬 계통발생)

-Consider all real matrices

-For non-square matrices (where n != m), the SBD always exists

-Focusing on square matrices, the determinant informs us whether a square matrix possesses an inverse matrix

-If the square nxn matrix possesses n linearly independent eigenvectors, then the matrix is non-defective and an eigen decomposition exists

-Repeated eigenvalues may result in defective matrices, which cannot be diagonalized

-Non-singular and non-defective matrices are not the same

Normal if Atranspose*A = A*Atranspose

Orthogonal if Atranspose*A = A*Atranspose = I

Normal matrices have a frequently encountered subset, the symmetric matrices $\boldsymbol{S} \in \mathbb{R}^{n \times n}$, which satisfy $\boldsymbol{S} = \boldsymbol{S}^\top$. Symmetric matrices have only real eigenvalues. A subset of the symmetric matrices consists of the positive definite matrices $\boldsymbol{P}$ that satisfy the condition of $\boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$. In this case, a unique *Cholesky decomposition* exists (Theorem 4.18). Positive definite matrices have only positive eigenvalues and are always invertible (i.e., have a nonzero determinant).

Another subset of symmetric matrices consists of the *diagonal matrices* $\boldsymbol{D}$. Diagonal matrices are closed under multiplication and addition, but do not necessarily form a group (this is only the case if all diagonal entries are nonzero so that the matrix is invertible). A special diagonal matrix is the identity matrix $\boldsymbol{I}$.