Singular Value Decomposition

Singular Value Decomposition I

- ▼ Keywords
    - ▼ Singular value decomposition

SVD (Singular Value Decomposition)

1. SVD
    a. given a matrix A (m x n) where m > n
    b. SVD gives $A = U\Sigma V^T$
2. SVD as sum of outer products
    a. $A = U\Sigma V^T = \Sigma_{1 \le i \le n} \sigma_i u_i v_i^T$
        i. $\sigma_1 \ge \sigma_2 \ge ...\sigma_n$
3. Another perspective of SVD
    a. $Av_i = \sigma_i u_i (i = 1, ..., n)$
    b. $V^{-1} = V^T$
        i. V (n x n) has orthonormal columns
        ii. thus, $AV = U\Sigma(A = U\Sigma V^T)$

Singular Value Decomposition II

- ▼ Keywords
    - ▼ Spectral theorem
    - ▼ Symmetric matrix
    - ▼ Positive definite matrix

Computing SVD

1. Form AA^T (m x m) and A^TA (n x n) and compute eigendecomposition of each
    a. $AA^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma\Sigma^T U^T = U\Sigma^2 U^T$

    b. $A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T = V\Sigma^2V^T$

2. Find the following facts
    a. **orthogonal eigenvector matrices U and V**
    b. **eigenvalues in Σ^2 that are all positive**
    c. eigenvalues in Σ^2 that are shared by AA^T and A^TA
        i. AA^T and A^TA are symmetric positive (semi-)definite
            - Symmetric
                - $(AA^T)^T = AA^T$
                - $(A^TA)^T = A^TA$
            - Positive (semi-)definite
                - $x^TAA^Tx = (A^Tx)^T(A^Tx) = ||A^Tx||^2 \geq 0$
                - $x^TA^TAx = (Ax)^T(Ax) = ||Ax||^2 \geq 0$

Diagonalization of Symmetric Matrices

1. A (n x n) is diagonalizable if and only if n linearly independent eigenvectors exist
    a. $(AA^T)^T = AA^T$
2. symmetric matrix S (n x n) where $S^T = S$ is always diagonalizable
    a. S is orthogonally diagonalizable
    b. i.e., eigenvectors are not only linearly independent, but also orthogonal to each other

Spectral Theorem of Symmetric Matrices

1. Consider a symmetric matrix S (n x n) where $S^T = S$
2. S has n real eigenvalues, counting multiplicities
3. The dimension of the eigenspace for each eigenvalue equals the multiplicity of λ as a root of the characteristic equation
    a. det 각 근의 중근의 개수 (algebraic multiplicity)
    b. 이에 해당하는 eigenspace 의 basis 의 개수 (geometric multiplicity)
    c. 위의 두 multiplicity 의 개수가 똑같아야 max 값인 n개의 eigenvalues 를 구할 수 있음

4. The eigenspaces are mutually orthogonal
    a. i.e., eigenvectors corresponding to different eigenvalues are orthogonal
5. To sum up, S is orthogonally diagonalizable

Spectral Decomposition

1. Eigendecomposition of a symmetric matrix is known as spectral decomposition
    a. $S = UDU^{-1} = UDU^T = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + ... + \lambda_n u_n u_n^T$
        i. $\lambda_i u_j u_j^T$ can be viewed as a projection matrix onto the subspace spanned by u_j, scaled by its eigenvalue λ_i

Positive Definite Matrices

1. A (n x n) is positive definite if and only if the eigenvalues of A are all positive
    a. A (n x n) is positive definite if $x^T Ax > 0$
    b. A (n x n) is positive semi-definite if $x^T Ax \geq 0 (x \neq 0)$
2. Symmetric positive definite matrices
    a. if S (n x n) is symmetric and positive-definite, then the spectral decomposition will have all positive eigenvalues
        i. $S = UDU^T = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + ... + \lambda_n u_n u_n^T$ where $\lambda_j > 0 (j = 1, ..., n)$

Things to Note

1. Given any rectangular matrix A (m x n), its SVD always exists
2. Given a square matrix A (n x n), its eigendecomposition does not always exist, but its SVD always exists
3. Given a square, symmetric positive (semi-)definite matrix S (n x n), its eigendecomposition always exists, and it is actually the same as its SVD

Eigen Decomposition & Singular Value Decomposition in ML

▼ Keywords

  ▼ Principal component analysis

▼ Gram matrix

▼ Low-rank approximation

▼ Dimension-reducing transformation

Eigendecomposition in Machine Learning

1. In machine learning, usually handle symmetric positive (semi-)definite matrix
2. Given a (feature-by-data item) matrix A (m x n)
3. A^TA represents a (data item-by-data item) similarity matrix between all paris of data items, where the similarity is computed as an inner product
    a. correlation 값이 높다 == inner product 값이 크다
4. AA^T represents a (feature-by-feature) similarity matrix between all pairs of features, indicating a kind of correlations between features
    a. covariance matrix in principal component analysis
    b. gram matrix in style transfer

Low-Rank Approximation of a Matrix

1. SVD of a rectangular matrix A (m x n) can be represented as the sum of outer products
    a. $A = U\Sigma V^T = \Sigma_{1 \leq i \leq n} \sigma_i u_i v_i^T$
2. The problem of the best low-rank approximation
    a. $\hat{A}_r = argmin_{A_r} ||A - A_r||_F$ subject to $rank(A_r) \leq r$
        i. F는 frobenius norm 을 의미함
        ii. norm 을 matrix 단위로 확장한 것으로, 모든 matrix 의 element 를 제곱해서 더한 값이 됨
3. The optimal solution is given as
    a. $\hat{A}_r = \Sigma_{1 \leq i \leq r} \sigma_i u_i v_i^T$
4. Approximate A as A_r

Dimension-Reducing Transformation

1. Given a (feature-by-data item) matrix X (m x n)
2. Consider the linear transformation, $G^T : x \rightarrow y$

a. 최적의 솔루션은 orthonormal 한 세 개의 projection vector

b. 이 벡터들은 SVD에서 u_i들을 모아서 row vector로 만든 matrix 가 됨

3. Goal

a. pairwise similarity matrix 정보를 가장 잘 보존하도록 하는 차원 축소된 버전의 표현형을 얻는 것

b. $\hat{G} = argmin_G ||S - X^T G G^T X||_F$ subject to $G^T G = I_K$

   a. given $X = U \Sigma V^T = \Sigma_{1 \leq i \leq n} \sigma_i u_i v_i^T$

   b. optimal solution is $\hat{G} = U_r$