Module 6: Dimensionality Reduction

Lecture 29: Singular Value Decomposition (SVD)

Prev topic Next topic Prev page
Next page

The Lecture Contains:

Singular Value Decomposition (SVD)

- Factorization of a matrix

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

- If A is of size $m \times n$, then U is $m \times m$, V is $n \times n$ and Σ is $m \times n$ matrix
- Columns of U are eigenvectors of AA^T
  - Left singular vectors
  - $UU^T = I_m$ (orthonormal)
- Columns of V are eigenvectors of $A^T A$
  - Right singular vectors
  - $V^T V = I_n$ (orthonormal)
- σ_{ii} are the singular values
  - Σ is diagonal
  - Singular values are positive squareroots of eigenvalues of AA^T or $A^T A$
- $\sigma_{11} \geq \sigma_{22} \geq \cdots \geq \sigma_{nn}$ (assuming n singular values)

Prev topic Next topic Prev page
Next page

SVD of real symmetric matrix

- A is real symmetric of size $n \times n$
- $A = A^T$
- $U = V$ since $A^T A = AA^T = A^2$

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

- Q is of size $n \times n$ and contains eigenvectors of A^2
- This is called spectral decomposition of A
- Σ contains n singular values
- Eigenvectors of A = eigenvectors of A^2
- Eigenvalues of A = squareroot of eigenvalues of A^2
- Eigenvalues of A = singular values of A

Transformation using SVD

- Transformed data

$$T = AV = U\Sigma$$

- V is called SVD transform matrix
- Essentially, T is just a rotation of A
- Dimensionality of T is n
- n different basis vectors than the original space
- Columns of V give the basis vectors in rotated space

Prev topic Next topic Prev page
Next page

Transformation using SVD

- Transformed data

$$T = AV = U\Sigma$$

- V is called SVD transform matrix
- Essentially, T is just a rotation of A
- Dimensionality of T is n
- n different basis vectors than the original space
- Columns of V give the basis vectors in rotated space
- V shows how each dimension can be represented as a linear combination of other dimensions
    - Columns are input basis vectors
- U shows how each object can be represented as a linear combination of other objects
    - Columns are output basis vectors

Transformation using SVD

- Transformed data

$$T = AV = U\Sigma$$

- V is called SVD transform matrix
- Essentially, T is just a rotation of A
- Dimensionality of T is n
- n different basis vectors than the original space
- Columns of V give the basis vectors in rotated space
- V shows how each dimension can be represented as a linear combination of other dimensions
    - Columns are input basis vectors
- U shows how each object can be represented as a linear combination of other objects
    - Columns are output basis vectors
- Lengths of vectors are preserved

$$||\vec{a_i}||_2 = ||\vec{t_i}||_2$$

Prev topic Next topic Prev page
Next page

Example

$$A \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{bmatrix} = U \begin{bmatrix} -0.80 & 0.22 & 0.05 & 0.54 \\ -0.56 & -0.20 & -0.34 & -0.71 \\ -0.16 & -0.31 & 0.90 & -0.21 \\ -0.01 & -0.89 & -0.22 & 0.37 \end{bmatrix}$$

$$\times \Sigma \begin{bmatrix} 5.54 & 0 \\ 0 & 1.24 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \times V^T \begin{bmatrix} -0.39 & -0.92 \\ 0.92 & -0.39 \end{bmatrix}^T$$

$$T = AV = U\Sigma = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{bmatrix} \times \begin{bmatrix} -0.39 & -0.92 \\ 0.92 & -0.39 \end{bmatrix} = \begin{bmatrix} -4.46 & 0.27 \\ -3.15 & -0.25 \\ -0.92 & -0.39 \\ -0.06 & -1.15 \end{bmatrix}$$

Prev topic Next topic Prev page
Next page

Example: compact form

$$A\begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{bmatrix} = U\begin{bmatrix} -0.80 & 0.22 \\ -0.56 & -0.20 \\ -0.16 & -0.31 \\ -0.01 & -0.89 \end{bmatrix}$$

$$\times \Sigma\begin{bmatrix} 5.54 & 0 \\ 0 & 1.24 \end{bmatrix} \times V^T\begin{bmatrix} -0.39 & -0.92 \\ 0.92 & -0.39 \end{bmatrix}^T$$

- If A is of size $m \times n$, then U is $m \times n$, V is $n \times n$ and Σ is $n \times n$ matrix
- Works because there at most n non-zero singular values in Σ

Dimensionality reduction using SVD

$$A = U\Sigma V^T = \sum_{i=1}^{n}(u_i \sigma_{ii} v_i^T)$$

- Use only k dimensions
- Retain first k columns for U and V and first k values for Σ
- First k columns of V give the basis vectors in reduced space
- Best rank k approximation in terms of sum squared error

$$A \approx \sum_{i=1}^{k}(u_i \sigma_{ii} v_i^T) = U_{1\ldots k}\Sigma_{1\ldots k}V_{1\ldots k}^T$$

$$T \approx AV_{1\ldots k}$$

Prev topic Next topic Prev page
Next page

Example of dimensionality reduction (k = 1)

$$A \approx A_k = U_k \begin{bmatrix} -0.80 & 0 & 0 & 0 \\ -0.56 & 0 & 0 & 0 \\ -0.16 & 0 & 0 & 0 \\ -0.01 & 0 & 0 & 0 \end{bmatrix} \times \Sigma_k \begin{bmatrix} 5.54 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \times V_k^T \begin{bmatrix} -0.39 & 0 \\ -0.92 & 0 \end{bmatrix}^T$$

$$= \begin{bmatrix} 1.74 & 4.10 \\ 1.23 & 2.90 \\ 0.35 & 0.84 \\ 0.02 & 0.06 \end{bmatrix}$$

$$T \approx T_k = AV_k = U_k \Sigma_k = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{bmatrix} \times \begin{bmatrix} -0.39 & 0 \\ 0.92 & 0 \end{bmatrix} = \begin{bmatrix} -4.46 & 0 \\ -3.15 & 0 \\ -0.92 & 0 \\ -0.06 & 0 \end{bmatrix}$$

Prev topic Next topic Prev page
Next page

Compact way of dimensionality reduction (k = 1)

$$A \approx A_k = U_k \begin{bmatrix} -0.80 \\ -0.56 \\ -0.16 \\ -0.01 \end{bmatrix} \times \Sigma_k \begin{bmatrix} 5.54 \end{bmatrix} \times V_k^T \begin{bmatrix} -0.39 \\ -0.92 \end{bmatrix}^T$$

$$= \begin{bmatrix} 1.74 & 4.10 \\ 1.23 & 2.90 \\ 0.35 & 0.84 \\ 0.02 & 0.06 \end{bmatrix}$$

$$T \approx T_k = AV_k = U_k\Sigma_k = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 1 \\ -1 & 0.5 \end{bmatrix} \times \begin{bmatrix} -0.39 \\ 0.92 \end{bmatrix} = \begin{bmatrix} -4.46 \\ -3.15 \\ -0.92 \\ -0.06 \end{bmatrix}$$

Prev topic Next topic Prev page
Next page

How many dimensions to retain?

Module 6: Dimensionality Reduction

Lecture 29: Singular Value Decomposition (SVD)

How many dimensions to retain?

- There is no easy answer

How many dimensions to retain?

- There is no easy answer
- Concept of energy of a dataset
- Total energy is sum of squares of singular values (aka spread or variance)

$$E = \sum_{i=1}^{n} \sigma_{ii}^2$$

- Retain k dimensions such that $p\%$ of the energy is retained

$$E_k = \sum_{i=1}^{k} \sigma_{ii}^2$$

$$E_k / E \geq p$$

- Generally, p is between 80% to 95%

Module 6: Dimensionality Reduction

Lecture 29: Singular Value Decomposition (SVD)

Prev topic Next topic Prev page
Next page

How many dimensions to retain?

- There is no easy answer
- Concept of energy of a dataset
- Total energy is sum of squares of singular values (aka spread or variance)

$$E = \sum_{i=1}^{n} \sigma_{ii}^2$$

- Retain k dimensions such that $p\%$ of the energy is retained

$$E_k = \sum_{i=1}^{k} \sigma_{ii}^2$$

$$E_k / E \geq p$$

- Generally, p is between 80% to 95%
- In the above example, k = 1 retains 95.22% of the energy

Module 6: Dimensionality Reduction

Lecture 29: Singular Value Decomposition (SVD)

Prev topic Next topic Prev page
Next page

How many dimensions to retain?

- There is no easy answer
- Concept of energy of a dataset
- Total energy is sum of squares of singular values (aka spread or variance)

$$E = \sum_{i=1}^{n} \sigma_{ii}^2$$

- Retain k dimensions such that $p\%$ of the energy is retained

$$E_k = \sum_{i=1}^{k} \sigma_{ii}^2$$

$$E_k / E \geq p$$

- Generally, p is between 80% to 95%
- In the above example, k = 1 retains 95.22% of the energy
- Running time: $O(mnr)$ for A of size $m \times n$ and rank r