Lecture 5 Singular value decomposition

Weinan E[1,2] and Tiejun Li[2]

[1] Department of Mathematics,
Princeton University,
weinan@princeton.edu

[2] School of Mathematical Sciences,
Peking University,
tieli@pku.edu.cn
No.1 Science Building, 1575

Outline

Review and applications

QR for symmetric matrix

Numerical SVD

Singular value decomposition

Theorem (Singular value decomposition)

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then there exist $\boldsymbol{U} \in \mathbb{R}^{m\times m}$, $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ such that

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}$$

where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{m\times n}$. r is the rank of $\boldsymbol{A}$, $\sigma_i > 0$ are called singular values of $\boldsymbol{A}$, $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}, \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$ are orthogonal matrices.

It is straightforward that

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V} = \boldsymbol{V}^T\text{diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0)\boldsymbol{V}$$

i.e. the singular value $\sigma_i = \sqrt{\lambda_i(\boldsymbol{A}^T\boldsymbol{A})}$. Similarly we have $\sigma_i = \sqrt{\lambda_i(\boldsymbol{A}\boldsymbol{A}^T)}$.

About singular values

- To find the orthogonal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ is equivalent to find the eigenvectors of matrices $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^T$.
- If $\boldsymbol{A}$ is symmetric, the singular value matrix $\boldsymbol{\Sigma} = \boldsymbol{D}$, where $\boldsymbol{D} = diag(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$. λ_i is the eigenvalues of $\boldsymbol{A}$, and $\boldsymbol{V} = \boldsymbol{U}^T$.
- The 2-norm of a matrix

$$\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}(\boldsymbol{A}^T\boldsymbol{A})} = \sigma_{\max}.$$

- The 2-condition number

$$Cond_2(\boldsymbol{A}) = \|\boldsymbol{A}\|_2\|\boldsymbol{A}^{-1}\|_2 = \frac{\sigma_{\max}}{\sigma_{\min}}.$$

Generalized inverse of a matrix

- In general, if $\boldsymbol{A}$ is singular, $\boldsymbol{A}^{-1}$ doesn't exist! If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, there is no definition for $\boldsymbol{A}^{-1}$.
- We define the Moore-Penrose generalized inverse of $\boldsymbol{A}$ as

$$\boldsymbol{A}^{+} = \boldsymbol{V}^{T}\text{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)\boldsymbol{U}^{T}$$

for arbitrary matrix $\boldsymbol{A}$!

Least square problems

- Least square problem 1: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ may have more than one solution. If it has more than one solution we wish to pick one with $\|\boldsymbol{x}\|_2$ is the smallest, i.e., to find $\boldsymbol{x} \in \mathcal{S} = \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ such that

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_2$$

- Least square problem 2: if it has no solution we wish to pick one which is the solution of the following minimization problem

$$\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$$

- In any case we have the following solution by generalized inverse

$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{b}.$$

Multivariate linear regression

- Formulation

  Suppose we have a list of experimental data for a multi-variate function $Y = f(x_1, x_2, \dots, x_m)$, after taking the zero-th and first order terms, we approximate Y as

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

  The problem is how to recover β_i from the data?

- Naively consider the linear system

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}$$

  and $i = 1, \dots, n$. It may have no solution or have infinite solutions. This is reduced to the least square problem for

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{Y}$$

Multivariate linear regression

- We have

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

- Least square solution

$$\boldsymbol{\beta} = \boldsymbol{X}^{+}\boldsymbol{Y}$$

Principal component analysis (PCA)

- Object: For a multi-component problem, is it possible to catch very few but very important characters to reduce the scale or dimension of the problem?
- Answer: Yes! PCA can do this job!

Principal component analysis (PCA)

- PCA

  Suppose we have experimental data to n characters(特征) of t units(单元) for a biological species, which can be proposed a matrix under experiments or investigations as

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ y_{t1} & y_{t2} & \cdots & y_{tn} \end{pmatrix}$$

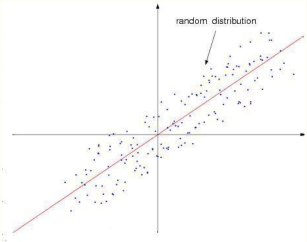
- Object: Intuitively, PCA is to find vectors $\boldsymbol{a}_i = (a_{1i}, a_{2i}, a_{ni})$ $(i = 1, \ldots, n)$ such that

$$\boldsymbol{F}_i = a_{1i}\boldsymbol{y}_1 + a_{2i}\boldsymbol{y}_2 + \cdots + a_{ni}\boldsymbol{y}_n, \quad i = 1, \ldots, n$$

  are perpendicular each other, and pick up some large components among $\|\boldsymbol{F}_i\|_2$. The analysis of $\boldsymbol{a}_i$ will give the main components of the problem.

Principal component analysis (PCA)

A geometrical interpretation of PCA for 2D coordinates analysis



- A mathematical rigorous interpretation (Projection maximization)

$$\max_{\|\boldsymbol{a}\|_2=1} \sum_{i=1}^{N} (\boldsymbol{x}_i \cdot \boldsymbol{a})^2 = \boldsymbol{a}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{a}$$

- Courant-Fisher's theorem gives PCA.

Principal component analysis (PCA)

- Step 1: non-dimensionalization

  Calculate the mean $\bar{y}_j = \dfrac{1}{t}\sum_{k=1}^{t} y_{kj}, \quad j = 1, 2 \ldots, n$

  Calculate variance $d_j = \sqrt{\sum_{k=1}^{t}(y_{kj} - \bar{y}_j)^2}, \quad j = 1, 2 \ldots, n$

  Transformation $x_{ij} = \dfrac{y_{ij} - \bar{y}_j}{d_j}, \quad i = 1, 2 \ldots, t; \;\; j = 1, 2 \ldots, n$

Non-dimensionalization is used to eliminate the effect of choice of unit (单位).

Principal component analysis (PCA)

- Step 2: Define principal component vector as

$$\boldsymbol{F}_i = a_{1i}\boldsymbol{x}_1 + a_{2i}\boldsymbol{x}_2 + \cdots + a_{ni}\boldsymbol{x}_n, \quad i = 1, \ldots, n$$

where $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{ti})$. In order the vectors are independent each other, we need

$$\boldsymbol{F}_i^T \boldsymbol{F}_j = 0, \quad i \neq j$$

i.e.

$$\boldsymbol{F}_i^T \boldsymbol{F}_j = (a_{1i}\ a_{2i}\ \ldots\ a_{ni})\boldsymbol{X}^T\boldsymbol{X}\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} = 0$$

Principal component analysis (PCA)

- Step 3: There exists orthogonal matrix $\boldsymbol{A}$ such that

$$\boldsymbol{A}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{A} = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

  and $\lambda_k \geq 0$ $(k = 1, \ldots, n)$. We have if $i \neq j$, the vectors $\boldsymbol{a}_i, \boldsymbol{a}_j$ in the i-th and j-th column will satisfy the independent condition, and

$$\|\boldsymbol{F}_i\|_2 = \lambda_i$$

- Step 4: Take the eigenvectors $\boldsymbol{a}_i$ corresponding to the first m biggest eigenvalues $(\lambda_1 > \lambda_2 > \cdots > \lambda_m > \cdots)$, and make linear combination

$$\boldsymbol{F}_i = a_{1i}\boldsymbol{x}_1 + a_{2i}\boldsymbol{x}_2 + \cdots + a_{ni}\boldsymbol{x}_n, \quad i = 1, 2, \ldots, m$$

  We will obtain the first m principal component vectors.

PCA and SVD

- If $\boldsymbol{X}$ has SVD

$$\boldsymbol{X} = \boldsymbol{U\Sigma V}$$

then we have $\boldsymbol{A} = \boldsymbol{V}^T$, and

$$\boldsymbol{V X}^T \boldsymbol{X V}^T = \boldsymbol{\Sigma}^T \boldsymbol{\Sigma}$$

- To find the first m principal component vectors is equivalent to find the first m principal (biggest) singular value and corresponding right singular vectors.

Outline

Review and applications

QR for symmetric matrix

Numerical SVD

Tri-diagonalization of symmetric matrix

- First transform symmetric $\boldsymbol{A}$ into tri-diagonal matrix $\boldsymbol{T}$

$$\boldsymbol{T} = \begin{pmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_{n-1} \\ & & \beta_{n-1} & \alpha_n \end{pmatrix}$$

  by a sequence of Householder transformations.
- The transformation procedure is the same as that for upper Hessenburg form with symmetry argument.

Tri-diagonalization of symmetric matrix

- The approach is to apply Householder transformation to $\boldsymbol{A}$ column by column.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

- Suitably choose Householder matrix $\boldsymbol{H}_1$ such that

$$\boldsymbol{H}_1 \cdot \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix} = \begin{pmatrix} a'_{11} \\ a'_{21} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \boldsymbol{H}_1 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}'_1 \end{pmatrix}$$

Tri-diagonalization of symmetric matrix

- Now we have

$$A_1 = H_1 A H_1 = \begin{pmatrix} a'_{11} & a'_{12} & \cdots & 0 \\ a'_{21} & a'_{22} & \cdots & a'_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & a'_{n2} & \cdots & a'_{nn} \end{pmatrix}$$

  by symmetry of A and A_1.
- The next step is the same for upper Hesseburg form. Finally we have tridiagonal form T and T has the same eigenvalues as A.

Implicit shifted QR for symmetric tridiagonal matrix

- Now we have symmetric tridiagonal $\boldsymbol{T}$ with diagonal entries $\alpha_i(i = 1, \ldots, n)$ and off-diagonal entries $\beta_i(i = 1, \ldots, n-1)$, one shifted QR step is

$$\boldsymbol{T} - \mu \boldsymbol{I} = \boldsymbol{Q}\boldsymbol{R}$$

$$\hat{\boldsymbol{T}} = \boldsymbol{R}\boldsymbol{Q} + \mu \boldsymbol{I}$$

In fact

$$\hat{\boldsymbol{T}} = \boldsymbol{Q}^T \boldsymbol{T} \boldsymbol{Q}$$

If we can find $\boldsymbol{Q}, \hat{\boldsymbol{T}}$ directly, we doesn't need the intermediate steps.

- In fact

$$\boldsymbol{Q}^T \boldsymbol{T} \boldsymbol{Q} = \boldsymbol{Q}^T (\boldsymbol{Q}\boldsymbol{R} + \mu \boldsymbol{I}) \boldsymbol{Q} = \boldsymbol{R}\boldsymbol{Q} + \mu \boldsymbol{I} = \hat{\boldsymbol{T}}.$$

Implicit shifted QR for symmetric tridiagonal matrix

- Find Givens matrix $\boldsymbol{G}_1 = \boldsymbol{G}(1,2;\theta_1)$ such that

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix}^T \cdot \begin{pmatrix} \alpha_1 - \mu \\ \beta_1 \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix}$$

- Define

$$\boldsymbol{T}_1 = \boldsymbol{G}_1^T \boldsymbol{T} \boldsymbol{G}_1.$$

We have

$$\boldsymbol{T}_1 = \begin{pmatrix} * & * & {\color{red}*} & & \\ * & * & * & & \\ {\color{red}*} & * & * & \ddots & \\ & & \ddots & \ddots & * \\ & & & * & * \end{pmatrix}$$

- We should zero out the term $\color{red}*$. That only needs another Givens matrix $\boldsymbol{G}_2$ multiplication.

Implicit shifted QR for symmetric tridiagonal matrix

- We can find Givens matrix $\boldsymbol{G}_2 = \boldsymbol{G}(2,3;\theta_2)$ such that the term * would be zero out.
- Define

$$\boldsymbol{T}_2 = \boldsymbol{G}_2^T \boldsymbol{G}_1^T \boldsymbol{T} \boldsymbol{G}_1 \boldsymbol{G}_2$$

We have

$$\boldsymbol{T}_2 = \begin{pmatrix} * & * & & & \\ * & * & * & {\color{red}*} & \\ & * & * & \ddots & \\ & {\color{red}*} & \ddots & \ddots & * \\ & & & * & * \end{pmatrix}$$

- We should zero out the term * again. That needs a Givens matrix multiplication again.

Implicit shifted QR for symmetric tridiagonal matrix

- Sequentially we have

$$\boldsymbol{T}_{n-2} = \begin{pmatrix} * & * & & & \\ * & * & * & & \\ & * & * & \ddots & {\color{red}*} \\ & & \ddots & \ddots & * \\ & & {\color{red}*} & * & * \end{pmatrix}$$

- Finally we obtain

$$\hat{\boldsymbol{T}} = \begin{pmatrix} * & * & & & \\ * & * & * & & \\ & * & * & \ddots & \\ & & \ddots & \ddots & * \\ & & & * & * \end{pmatrix}$$

Implicit shifted QR for symmetric tridiagonal matrix

- Iterating for $\hat{\boldsymbol{T}}$ to obtain the next QR step!
- In general the shift is chosen as the famous Wilkinson's shift: If the submatrix of $\boldsymbol{T}$

$$\boldsymbol{S} = \begin{pmatrix} \alpha_{n-1} & \beta_{n-1} \\ \beta_{n-1} & \alpha_n \end{pmatrix}$$

then choose μ one of the eigenvalues of $\boldsymbol{S}$ which is more closer to α_n.

$$\mu = \alpha_n + \delta - \mathsf{sign}(\delta)\sqrt{\delta^2 + \beta_{n-1}^2}$$

and $\delta = \frac{\alpha_n + \alpha_{n-1}}{2}$.

- The convergence will be very fast with this shift.

Outline

Review and applications

QR for symmetric matrix

Numerical SVD

Implicit QR method for singular value computation

- First transform $\boldsymbol{A}$ into upper bidiagonal matrix $\boldsymbol{B}$

$$\boldsymbol{B} = \begin{pmatrix} d_1 & f_2 & & \\ & d_2 & \ddots & \\ & & \ddots & f_n \\ & & & d_n \end{pmatrix}$$

  by a sequence of Householder transformations

$$\boldsymbol{A} \xrightarrow{\boldsymbol{U}_1} \text{eliminate the first column} \xrightarrow{\boldsymbol{V}_1} \text{eliminate the first row} \xrightarrow{\cdots}$$

$$\xrightarrow{\boldsymbol{U}_n} \text{eliminate the n-th column} = \begin{pmatrix} \boldsymbol{B} \\ \boldsymbol{0} \end{pmatrix}$$

- $\boldsymbol{A}$ has the same singular values as $\boldsymbol{B}$.

Implicit QR method for singular value computation

- First transform $\boldsymbol{A}$ into upper bidiagonal matrix $\boldsymbol{B}$

$$\boldsymbol{B} = \begin{pmatrix} d_1 & f_2 & & \\ & d_2 & \ddots & \\ & & \ddots & f_n \\ & & & d_n \end{pmatrix}$$

  by a sequence of Householder transformations

$$\boldsymbol{A} \xrightarrow{\boldsymbol{U}_1} \text{eliminate the first column} \xrightarrow{\boldsymbol{V}_1} \text{eliminate the first row} \xrightarrow{\cdots}$$

$$\xrightarrow{\boldsymbol{U}_n} \text{eliminate the n-th column} = \begin{pmatrix} \boldsymbol{B} \\ \boldsymbol{0} \end{pmatrix}$$

- Now we have

$$\boldsymbol{U}_n \cdots \boldsymbol{U}_1 \boldsymbol{A} \boldsymbol{V}_1 \cdots \boldsymbol{V}_{n-1} = \begin{pmatrix} \boldsymbol{B} \\ \boldsymbol{0} \end{pmatrix}$$

Implicit shifted QR method for singular value computation

- Basic idea: Implicitly apply shifted QR method to symmetric tridiagonal matrix $\boldsymbol{B}^T\boldsymbol{B}$ but without forming it.
- Steps:
  - Determine the shift μ. This is equivalent to the shift step for $\boldsymbol{B}^T\boldsymbol{B}$.
    Wilkinson shift: set μ is the eigenvalue of

$$\begin{pmatrix} d_{n-1}^2 + f_{n-1}^2 & d_{n-1}f_n \\ d_{n-1}f_n & d_n^2 + f_n^2 \end{pmatrix}$$

    closer to $d_n^2 + f_n^2$ to make the convergence faster.
  - Find Givens matrix $\boldsymbol{G}_1 = \boldsymbol{G}(1,2;\theta)$ such that

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix}^T \cdot \begin{pmatrix} d_1^2 - \mu \\ d_1 f_2 \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix}$$

    and compute $\boldsymbol{B}\boldsymbol{G}_1$.
    This is equivalent to apply $\boldsymbol{G}_1$ step for $\boldsymbol{B}^T\boldsymbol{B}$.

Implicit shifted QR method for singular value computation

- We have

$$BG_1 = \begin{pmatrix} * & * & & & \\ {\color{red}*} & * & * & & \\ & & * & \ddots & \\ & & & \ddots & * \\ & & & & * \end{pmatrix}$$

so we should zero out the term $\color{red}*$. We want to find P_2 and G_2 such that $P_2(BG_1)G_2$ is bidiagonal and $G_2 e_1 = e_1$.

This is equivalent to apply G_2 step for $G_1^T B^T B G_1$.

Implicit shifted QR method for singular value computation

- It is not difficult to find $\boldsymbol{P}_2$ and $\boldsymbol{G}_2$ by Givens transformation and we have

$$\boldsymbol{P}_2\boldsymbol{B}\boldsymbol{G}_1\boldsymbol{G}_2 = \begin{pmatrix} * & * & & & \\ & * & * & & \\ & {\color{red}*} & * & \ddots & \\ & & & \ddots & * \\ & & & & * \end{pmatrix}$$

so we should zero out the term $\color{red}{*}$. We want to find $\boldsymbol{P}_3$ and $\boldsymbol{G}_3$ such that $\boldsymbol{P}_3\boldsymbol{P}_2\boldsymbol{B}\boldsymbol{G}_1\boldsymbol{G}_2\boldsymbol{G}_3$ is bidiagonal and $\boldsymbol{G}_3\boldsymbol{e}_i = \boldsymbol{e}_i,\ i = 1, 2$.
These steps should be repeated until $\boldsymbol{B}\boldsymbol{G}_1$ becomes bidiagonal! It is equivalent to find $\boldsymbol{G}_i$ steps for symmetric tridiagonal matrix.

Implicit shifted QR method for singular value computation

- Finally we have

$$\boldsymbol{P}_{n-1}\cdots\boldsymbol{P}_2\boldsymbol{B}\boldsymbol{G}_1\cdots\boldsymbol{G}_{n-1} = \begin{pmatrix} * & * & & & \\ & * & * & & \\ & & * & \ddots & \\ & & & \ddots & * \\ & & & & * \end{pmatrix}$$

- Iterate until the off-diagonal entries converge to 0, and the diagonal entries converge to singular values!