Lecture 14: Least squares and Gram-Schmidt (Section 3.3-3.4)

Thang Huynh, UC San Diego

2/12/2018

Least squares

In practice,

$$A\boldsymbol{x} \approx \boldsymbol{b}.$$

Least squares

In practice,

$$A\boldsymbol{x} \approx \boldsymbol{b}.$$

- Definition. $\hat{\boldsymbol{x}}$ is a **least squares solution** of the system $A\boldsymbol{x} = \boldsymbol{b}$ if $\hat{\boldsymbol{x}}$ is such that $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible.

Theorem $\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ (the normal equation).

Application: least squares lines

- Example. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2, 1), (5, 2), (7, 3), (8, 3)$.

Application: least squares lines

- Example. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2,1), (5,2), (7,3), (8,3)$.
- Solution. The equations $y_i = \beta_1 + \beta_2 x_i$ in matrix form

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } y} .$$

Application: least squares lines

We need to find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

Application: least squares lines

We need to find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

We need to find the line $y = \widehat{\beta}_1 + \widehat{\beta}_2 x$, i.e. $(\widehat{\beta}_1, \widehat{\beta}_2)$, which minimizes the squared error

$$E^2 = \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 = (1 - \beta_1 - 2\beta_2)^2 + \ldots + (3 - \beta_1 - 8\beta_2)^2.$$

Application: least squares lines

We need to find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

We need to find the line $y = \widehat{\beta}_1 + \widehat{\beta}_2 x$, i.e. $(\widehat{\beta}_1, \widehat{\beta}_2)$, which minimizes the squared error

$$E^2 = \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 = (1 - \beta_1 - 2\beta_2)^2 + \ldots + (3 - \beta_1 - 8\beta_2)^2.$$

$\Longrightarrow$ least squares solution.

Application: least squares lines

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix},$$

and

$$X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix} \Longrightarrow \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}.$$

The least squares line is $y = \frac{2}{7} + \frac{5}{14}x$.

Other curves

We can also fit the experimental data (x_i, y_i) using other curves.

- Example. $y_i \approx \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ with parameters $\beta_1, \beta_2, \beta_3$.

Other curves

We can also fit the experimental data (x_i, y_i) using other curves.

- Example. $y_i \approx \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ with parameters $\beta_1, \beta_2, \beta_3$.

The equations $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ in matrix form:

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}.$$

Given the data (x_i, y_i), we then find the least squares solution to $X\boldsymbol{\beta} = \boldsymbol{y}$.

Gram-Schmidt

- Vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are orthonormal if

$$\boldsymbol{v}_i^T \boldsymbol{v}_j = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$$

- Gram-Schmidt orthonormalization:
  - Input: basis $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ for V
  - Output: orthonormal basis $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ for V.

Gram-Schmidt

$$
\begin{aligned}
\boldsymbol{b}_1 &= \boldsymbol{a}_1, & \boldsymbol{q}_1 &= \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} \\
\boldsymbol{b}_2 &= \boldsymbol{a}_2 - \langle \boldsymbol{a}_2, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1, & \boldsymbol{q}_2 &= \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|} \\
\boldsymbol{b}_3 &= \boldsymbol{a}_3 - \langle \boldsymbol{a}_3, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1 - \langle \boldsymbol{a}_3, \boldsymbol{q}_2 \rangle \boldsymbol{q}_2, & \boldsymbol{q}_3 &= \frac{\boldsymbol{b}_3}{\|\boldsymbol{b}_3\|} \\
&\vdots
\end{aligned}
$$

Gram-Schmidt

$$
\begin{aligned}
\boldsymbol{b}_1 &= \boldsymbol{a}_1, & \boldsymbol{q}_1 &= \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} \\
\boldsymbol{b}_2 &= \boldsymbol{a}_2 - \langle \boldsymbol{a}_2, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1, & \boldsymbol{q}_2 &= \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|} \\
\boldsymbol{b}_3 &= \boldsymbol{a}_3 - \langle \boldsymbol{a}_3, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1 - \langle \boldsymbol{a}_3, \boldsymbol{q}_2 \rangle \boldsymbol{q}_2, & \boldsymbol{q}_3 &= \frac{\boldsymbol{b}_3}{\|\boldsymbol{b}_3\|} \\
&\vdots
\end{aligned}
$$

- Example. Apply Gram-Schmidt to the vectors $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Gram-Schmidt

▶ Example. Apply Gram-Schmidt to the vectors $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Gram-Schmidt

- Example. Apply Gram-Schmidt to the vectors $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.
- Solution. $\frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \frac{1}{3}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$.

Theorem. The columns of an $m \times n$ matrix Q are orthonormal if and only if

$$Q^T Q = I_{n\times n}.$$

Proof.

Gram-Schmidt

Theorem. The columns of an $m \times n$ matrix Q are orthonormal if and only if

$$Q^T Q = I_{n\times n}.$$

Proof.

▶ Definition. An **orthogonal matrix** is a square matrix Q with orthonormal columns.

Gram-Schmidt

Theorem. The columns of an $m \times n$ matrix Q are orthonormal if and only if

$$Q^T Q = I_{n \times n}.$$

Proof.

- Definition. An **orthogonal matrix** is a square matrix Q with orthonormal columns.
- Example. $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is orthogonal.

Theorem. The columns of an $m \times n$ matrix Q are orthonormal if and only if

$$Q^T Q = I_{n \times n}.$$

Proof.

- **Definition.** An **orthogonal matrix** is a square matrix Q with orthonormal columns.
- **Example.** $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is orthogonal.
- **Example.** $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal.

Theorem. The columns of an $m \times n$ matrix Q are orthonormal if and only if

$$Q^T Q = I_{n \times n}.$$

Proof.

- **Definition.** An **orthogonal matrix** is a square matrix Q with orthonormal columns.
- **Example.** $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is orthogonal.
- **Example.** $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal.
- **Example.** Is $H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ orthogonal?

The QR decomposition

Let A be an $m \times n$ matrix of rank n. Then we have the **QR decomposition** $A = QR$,

- where Q is $m \times n$ and has orthonormal columns, and
- R is upper triangular, $n \times n$ and invertible.

The QR decomposition

Let A be an $m \times n$ matrix of rank n. Then we have the **QR decomposition** $A = QR$,

- where Q is $m \times n$ and has orthonormal columns, and
- R is upper triangular, $n \times n$ and invertible.

▶ Example. Find the QR decomposition of $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$.

▶ $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $R = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$.

The QR decomposition

Let A be an $m \times n$ matrix of rank n. Then we have the **QR decomposition** $A = QR$,

- where Q is $m \times n$ and has orthonormal columns, and
- R is upper triangular, $n \times n$ and invertible.

▶ Example. Find the QR decomposition of $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$.

▶ $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $R = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$.

In general, to obtain $A = QR$

- Gram-Schmidt on (columns of) A, to get Q.
- Then $R = Q^T A$.