Chapter 6. Linear Least Squares Data Fitting

Section 1. Normal equation

Consider the $m \times n$ liner system of equations,

$$A\mathbf{x} = \mathbf{b}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

When $m > n$ the system is overdetermined, and in general has no solutions. Let

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}$$

be the **remainder**. Since $\mathbf{r} \neq 0$ in general, an alternative way is to find $\mathbf{x}$ such that $\mathbf{r}$ is minimized in norm, i.e.,

$$\|\mathbf{r}\|_2^2 = \|\mathbf{b} - A\mathbf{x}\|_2^2$$

is minimized. This solution $\mathbf{x}$ is called a **least square solution**.

Theorem 6.1. (existence and uniqueness) The above linear least square problem always has solutions. If $\text{null}(A) = 0$, then the solution is unique. ∎

In the following, we assume that the least square solution is unique, i.e., we assume that the column vectors of A are linearly independent.

Theorem 6.2. (normal equation) Let $\mathbf{x}$ be the least square solution, then the remainder $\mathbf{r}$ satisfies

$$A^t\mathbf{r} = 0$$

or equivalently,

$$A^tA\mathbf{x} = A^t\mathbf{b}$$

This equation is called the normal equation, it is a $n \times n$ linear system. Since we assumed that the columns of A are linearly independent, A^tA is nonsingular and positive definite. Therefore, any method for solving square linear systems can be applied to solve this system. However, when the condition number of A is large, solving the normal equation directly is not efficient. Consider the special case, $m = n$, and consider the condition number of A^tA

$$\begin{aligned} \text{cond}(A^tA) &= \left\|A^tA\right\|_2 \cdot \left\|(A^tA)^{-1}\right\|_2 \\ &= \sqrt{\lambda((A^tA)^t(A^tA))} \cdot \sqrt{\lambda(((A^tA)^{-1})^t((A^tA)^{-1})} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\lambda((A^tA)^2)} \cdot \sqrt{\lambda((A^tA)^{-1})^2} \\
&= \lambda(A^tA) \cdot \lambda((A^tA)^{-1}) \\
&= [\mathrm{cond}(A)]^2
\end{aligned}
$$

where $\lambda(A)$ denotes the largest eigenvalue of A. Thus, the condition of A^tA can be very large if the condition of A is large.

Section 2. QR Factorization

In this section, we discuss the QR factorization method for solving the linear least square problems.

Theorem 6.3. **(QR factorization)** Suppose the columns of the $m \times n$ matrix A are linearly independent, then A has the QR factorization,

$$A = QR$$

where $Q_{m\times n}$ has orthogonal columns, and $R_{n\times n}$ is an upper triangle matrix. If we restrict the sign of the diagonal entries of R, the factorization is unique.

The matrices Q and R can be computed step by step as follows. Let

$$
\begin{aligned}
A &= [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n], \\
Q &= [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n], \\
R &= \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ & r_{22} & r_{23} & \cdots & r_{2n} \\ & & r_{33} & \cdots & r_{3n} \\ & & & \ddots & \vdots \\ & & & & r_{nn} \end{bmatrix}
\end{aligned}
$$

Let the diagonal entries of R be positive. Then from $A = QR$ we have

$$
\begin{aligned}
&\mathbf{a}_1 = r_{11}\mathbf{q}_1, \\
&\mathbf{a}_2 = r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2, \\
&\mathbf{a}_3 = r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3, \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\mathbf{a}_n = r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + \cdots + r_{nn}\mathbf{q}_n.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&r_{11} = \|\mathbf{a}_1\|_2, \mathbf{q}_1 = \mathbf{a}_1/r_{11}, \\
&r_{12} = \mathbf{q}_1^t\mathbf{a}_2, r_{22} = \|\mathbf{a}_2 - r_{12}\mathbf{q}_1\|_2, \mathbf{q}_2 = (\mathbf{a}_2 - r_{12}\mathbf{q}_1)/r_{22}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&r_{ik} = \mathbf{q}_i^t\mathbf{a}_k, i = 1, 2, \cdots, k-1, r_{kk} = \left\|\mathbf{a}_k - \sum_{i=1}^{k-1} r_{ik}\mathbf{q}_i\right\|_2, \mathbf{q}_k = \left(\mathbf{a}_k - \sum_{i=1}^{k-1} r_{ik}\mathbf{q}_i\right)/r_{kk}
\end{aligned}
$$

This is the Gram-Schmidt orthogonal process. However, this algorithm may not be stable due to round-off errors. To make it stable, we can normalize the vector $\mathbf{a}_i$ at each step.

Once we have the QR factorization, the least square solution is

$$\mathbf{x} = (A^tA)A^t\mathbf{b} = (R^tQ^tQR)^{-1}R^tQ^t\mathbf{b} = R^{-1}Q^t\mathbf{b}$$

Let

$$Q^t\mathbf{b} = \mathbf{c}$$

Then,

$$R\mathbf{x} = \mathbf{c}$$

This system can be solved easily, since R is an upper triangular matrix.

Section 3. Householder Transformation

To computer the QR factorization, we left multiply A by successive orthogonal matrices to obtain an upper triangular matrix R

$$R = H_nH_{n-1}\cdots H_1A$$

where $H_i, i = 1, \cdots, n$ are $m \times m$ orthogonal matrices and has the form

$$H_i = I - 2\mathbf{u}_i\mathbf{u}_i^t$$

with $\mathbf{u}_i^t\mathbf{u}_i = 1$. Matrix of this form is called the Householder Transformation (or Householder Matrix). The Householder Transformation has the property that it can transform any vector $\mathbf{x}$ to another vertor which is parallel to a given unit vector $\mathbf{g}$ and has the same length with $\mathbf{x}$, i.e., if $H = I - 2\mathbf{uu}$ is a Householder matrix, then

$$H\mathbf{x} = ||\mathbf{x}||_2\mathbf{g}$$

In this case,

$$\mathbf{u} = \frac{\mathbf{x} - ||\mathbf{x}||_2\mathbf{g}}{||\mathbf{x} - ||\mathbf{x}||_2\mathbf{g}||_2}$$

Let

$$A = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n]$$

then

$$H_1A = [H_1\mathbf{a}_1, H_1\mathbf{a}_2, \cdots, H_1\mathbf{a}_n]$$

We require

$$H_1\mathbf{a}_1 = \alpha_1\mathbf{e}_1$$

where

$$\alpha_1 = -\text{sign}(a_{11})\sqrt{a_{11}^2 + a_{21}^2 + \cdots + a_{n1}^2}$$

Then,

$$\mathbf{u}_1 = \frac{\mathbf{a}_1 - \alpha_1 \mathbf{e}_1}{\|\mathbf{a}_1 - \alpha_1 \mathbf{e}_1\|_2}$$

The first column of $H_1 A$ are all zeros except the first entry. Similar idea is applied to the submatrix by eliminating the first row and first column of $H_1 A$, denoted by $\tilde{A}_2$. More specifically, let

$$H_2 = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{H}_2 \end{bmatrix}$$

where $\tilde{H}_2 = I_{m-1} - 2\mathbf{u}_2\mathbf{u}_2^t$. Then,

$$H_2 H_1 A = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{H}_2 \end{bmatrix} \begin{bmatrix} \alpha_1 & \mathbf{b}_1^t \\ 0 & \tilde{A}_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 & \mathbf{b}_1^t \\ 0 & \tilde{H}_2 \tilde{A}_2 \end{bmatrix}$$

Select $\mathbf{u}_2$ such that the first column of $\tilde{A}_2$ are all zeros except the first entry. This procedure is continued until the last column of A. Let

$$Q = H_n H_{n-1} \cdots H_1$$

Then Q is an orthogonal matrix, and

$$QA = R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

where R_1 is an $n \times n$ upper triangular matrix. If we let

$$Q^t = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$$

where Q_1 is an $m \times n$ matrix with columns orthogonal, then

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1$$

which is the QR factorization discussed in last section. Let

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}$$

then,

$$Q\mathbf{r} = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} - \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{c} - R_1\mathbf{x} \\ \mathbf{d} \end{bmatrix}$$

where

$$\begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} = Q\mathbf{b}$$

Since Q is orthogonal, we have

$$\|\mathbf{r}\|_2^2 = \|Q\mathbf{r}\|_2^2 = \|\mathbf{c} - R_1\mathbf{x}\|_2^2 + \|\mathbf{d}\|_2^2$$

Obviously, when $\mathbf{x}$ is the solution of

$$R_1\mathbf{x} = \mathbf{c}$$

$\mathbf{r}$ is minimized.