Mathematics of least squares

Walter Mudzimbabwe

Solving LS problems by method of Normal equations

One of the most common methods for solving LS is the method of Normal equations. It involves solving the system

$$A^T Ax = A^T b$$

If rank(A) $= n$ then the system is positive definite and has solution $x = x_{LS}$.
The following steps are used to solve an LS problem using Normal equations:

1. Calculate $C = A^T A$.
2. Calculate $d = A^T b$.
3. Compute the Cholesky factorisation of C i.e., find G such that $C = GG^T$.
4. From $GG^T x = A^T b$, let $y = G^T x$ so that $Gy = d$.
5. Solve for y in $Gy = d$ using forward substitution.
6. Solve for x in $G^T x = y$ using backward substitution.

Example: Solving LS problems by method of Normal equations

Solve LS problem with

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

using the method of Normal equations.
Solution: $\mathrm{rank}(A) = 2 = n$.

$$C = A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 12 \\ 12 & 56 \end{bmatrix}$$

$$d = A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 12 \end{bmatrix}$$

The Cholesky decomposition of C yields

$$G = \begin{bmatrix} 1.7321 & 0. \\ 6.9282 & 2.8284 \end{bmatrix}.$$

Now

$$\begin{aligned} Gy &= d \\ \Rightarrow \begin{bmatrix} 1.7321 & 0. \\ 6.9282 & 2.8284 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} &= \begin{bmatrix} 3 \\ 12 \end{bmatrix} \\ y_1 &= 3/1.73205081 = 1.7320 \\ y_2 &= (12 - 1.7320 * 6.9282)/2.8284 \\ &\approx 0 \end{aligned}$$

Finally,

$$
\begin{aligned}
G^T x &= y \\
\Rightarrow \begin{bmatrix} 1.7321 & 6.9282 \\ 0 & 2.8284 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 1.7321 \\ 0 \end{bmatrix} \\
x_1 &= 1 \\
x_2 &= 0
\end{aligned}
$$

Therefore $x_{LS} = (1, 0)^T$.

Solving LS problems by classical Gram-Schmidt (CGS) algorithm

If $\mathrm{A} = [\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_n] = \mathrm{QR}$ and $\mathrm{Q} = [\mathrm{q}_1, \mathrm{q}_2, \ldots, \mathrm{q}_m]$. From matrix multiplication, we know that

$$a_{pk} = \sum_{i=1}^{k} q_{pi} r_{ik}, \quad k = 1, 2, \ldots, n.$$

Therefore

$$\mathrm{a}_k = \sum_{i=1}^{k} r_{ik} \mathrm{q}_i, \quad k = 1, 2, \ldots, n. \tag{1}$$

By orthonormality of q_i, $\mathrm{q}_i^T \mathrm{q}_j = 0$ when $i \neq j$ and $\mathrm{q}_i^T \mathrm{q}_i = 1$.

Therefore

$$
\begin{aligned}
\mathrm{q}_i^T \mathrm{a}_k &= \mathrm{q}_i^T \sum_{j=1}^{k} r_{jk} \mathrm{q}_j \\
&= \sum_{j=1}^{k} r_{jk} \mathrm{q}_i^T \mathrm{q}_j \\
&= r_{ik} \mathrm{q}_i^T \mathrm{q}_i \\
&= r_{ik}, \text{ since } \mathrm{q}_i^T \mathrm{q}_i = 1
\end{aligned}
$$

From (1)

$$\mathrm{a}_k = \sum_{i=1}^{k-1} r_{ik}\mathrm{q}_i + r_{kk}\mathrm{q}_k$$

$$\text{rearranging,} \qquad \mathrm{q}_k = \frac{\mathrm{a}_k - \sum_{i=1}^{k-1} r_{ik}\mathrm{q}_i}{r_{kk}}$$

Let

$$\mathrm{z}_k = \mathrm{a}_k - \sum_{i=1}^{k-1} s_{ik}\mathrm{q}_i, \text{ where } s_{ik} = \mathrm{q}_i^T\mathrm{q}_k$$

and

$$r_{kk} = ||\mathrm{z}_k||_2^2 = \mathrm{z}_k^T\mathrm{z}_k.$$

Given $A \in \mathbb{R}^{m\times n}$ then the following is the classical Gram-Schmidt (CGS) algorithm which computes the decomposition $A = QR$:

$$
\begin{aligned}
&\text{For } k = 1, 2 \ldots, n \\
&\qquad s_{ik} = q_i^T q_k, \qquad i = 1, 2 \ldots, k-1 \\
&\qquad z_k = a_k - \sum_{i=1}^{k-1} s_{ik} q_i \\
&\qquad r_{kk} = ||z_k||_2^2 \\
&\qquad q_k = z_k / r_{kk} \\
&\qquad r_{ik} = s_{ik} / r_{kk}, \qquad i = 1, 2 \ldots, k-1
\end{aligned}
$$

Solving LS problems using QR decomposition

CGS computes $A = Q_1R_1, Q_1 \in \mathbb{R}^{m\times n}, R_1 \in \mathbb{R}^{n\times n}$ where $Q_1^TQ_1 = I_n$ and R_1 is upper triangular.

<u>Exercise:</u> Show that the normal equations $(A^TA)x = A^Tb$ become $R_1x = Q_1^Tb$ using CGS.

<u>Solution:</u>

$$\begin{aligned} A^TA &= (Q_1R_1)^TQ_1R_1 \\ &= R_1^TQ_1^TQ_1R_1 \\ &= R_1^TI_nR_1 \\ &= R_1^TR_1 \end{aligned}$$

Therefore

$$\begin{aligned} (A^TA)x = A^Tb \text{ implies } (R_1^TR_1)x &= R_1^TQ_1^Tb \\ \text{implies} \qquad R_1x &= Q_1^Tb \end{aligned}$$