Mathematics of least squares

Walter Mudzimbabwe

Least square problem

Consider the least squares (LS) problem:

Find $x \in \mathbb{R}^n$ such that $Ax = b$, $A \in \mathbb{R}^{m\times n}$, $(m > n)$ and $b \in \mathbb{R}^m$.

The system is overdetermined when there are more equations than unknowns, $m > n$.
Usually there is no solution for overdertemined systems.
So we rather solve the problem:

$$\min_{x\in\mathbb{R}^n} ||Ax - b||_2, \text{ where } A \in \mathbb{R}^{m\times n}, \ b \in \mathbb{R}^m.$$

Important properties of solution

1.) If $x^* \in \mathbb{R}^n$ solves

$$\min_{x \in \mathbb{R}^n} ||Ax - b||_2$$

i.e., $||Ax^* - b||_2$ is the minimum then

$$A^T(b - Ax^*) = 0.$$

2.) $x^* \in \mathbb{R}^n$ is not necessarily unique.

Exercise

Verify that

$$A^T(b - Ax^*) = 0$$

where x^* solves the least square problem for the matrices

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \quad x^* = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Solution: we consider the minimum using calculus of the function

$$f(x) = \frac{1}{2}||b - Ax||_2^2.$$

Therefore

$$f(x^*) = \frac{1}{2}[(b_1 - a_{11}x_1 - a_{12}x_2)^2 + (b_2 - a_{21}x_1 - a_{22}x_2)^2 + (b_3 - a_{31}x_1 - a_{32}x_2)^2]$$

Exercise

Therefore

$$\bigtriangledown f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$

$$= \begin{bmatrix} -a_{11}(b_1 - a_{11}x_1 + a_{12}x_2) - a_{21}(b_2 - a_{21}x_1 - a_{22}x_2) - a_{31}(b_3 - a_{31}x_1 - a_{32}x_2) \\ -a_{12}(b_1 - a_{11}x_1 - a_{22}x_2) - a_{22}(b_2 - a_{21}x_1 - a_{22}x_2) - a_{32}(b_3 - a_{31}x_1 - a_{32}x_2) \end{bmatrix}$$

$$= - \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix} \begin{bmatrix} (b_1 - a_{11}x_1 + a_{12}x_2) \\ (b_2 - a_{21}x_1 + a_{22}x_2) \\ (b_3 - a_{31}x_1 + a_{32}x_2) \end{bmatrix}$$

$$= -A^T(b - Ax^*)$$

We know that from calculus to miminise a $f(x)$ we equate it's derivative to zero and solve for x, i.e.,

$$A^T(b - Ax^*) = 0.$$

The system of equations $A^T(b - Ax^*) = 0$ is known as normal equations.

Solving LS using SVD

Let $A = \sum_{j=1}^{r} \sigma_j u_j v_j^T$ where $r = \text{rank}(A)$ and

$$U = [u_1, u_2, \cdots, \cdots, u_m] \in \mathbb{R}^{m \times m}$$
$$V = [v_1, v_2, \cdots, v_n] \in \mathbb{R}^{n \times n}$$

be SVD of $A \in \mathbb{R}^{m \times n}$, $m \geq n$. If $b \in \mathbb{R}^m$ then

$$x_{LS} = \sum_{j=1}^{r} \frac{1}{\sigma_j} u_j^T b v_j.$$

So if you know the SVD of A in an LS problem then you can solve the LS by using elements in the SVD decomposition.

Example: Solving LS using SVD

Use SVD method to solve LS problem with

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

given that the SVD of A is

$$U = \begin{bmatrix} -0.28 & 0.87 & 0.41 \\ -0.54 & 0.21 & -0.82 \\ -0.79 & -0.45 & 0.41 \end{bmatrix}, \quad V = \begin{bmatrix} -0.21 & 0.98 \\ -0.98 & -0.21 \end{bmatrix}$$

and singular values are 7.65 and 0.64.

Example: Solving LS using SVD

<u>Solution</u>:
Note that rank(A) = 2.
The solution of the LS is

$$
\begin{aligned}
x^* = \sum_{j=1}^{2} \frac{1}{\sigma_j} u_j^T b v_j &= \frac{1}{7.65} u_1^T b v_1 + \frac{1}{0.64} u_2^T b v_2 \\
&= \frac{1}{7.65}[-0.28, -0.54, -0.79] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -0.21 \\ -0.98 \end{bmatrix} \\
&\quad + \frac{1}{0.64}[0.87, 0.21, -0.45] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -0.98 \\ -0.21 \end{bmatrix} \\
&= \begin{bmatrix} -0.27 \\ -1.71 \end{bmatrix}.
\end{aligned}
$$