Overdetermined System of Linear Equations

In the previous lecture, we learned how to solve an n by n linear systems using both direct (Gauss-elimination, Gauss-Jordan, etc.) and iterative methods (Jacobi and Gauss-Seidel).

Today, we are going to discuss solving an $m \times n$ linear system

$$Ax = b, \quad A \in \mathbb{R}^{m \times n}$$

where $m > n$, that is

- Numbers of rows in matrix A is larger than the number of columns
- Number of equations are larger than the number of variables

Example 1.

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 - x_2 &= 3 \\ x_1 + 2x_2 &= 2 \end{aligned}$$

Graph each equation in $x_1 x_2$ plane, and check whether there exists any single intersection point.

Example 1.

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 - x_2 &= 3 \\ x_1 + 2x_2 &= 2 \end{aligned}$$

Graph each equation in $x_1 x_2$ plane, and check whether there exists any single intersection point.
We observe that there is no exact solution. Can we find one point that roughly satisfy in all equation?

$$\min_{x_1, x_2} \quad (1 - x_1 - x_2)^2 + (3 - x_1 + x_2)^2 + (2 - x_1 - 2x_2)^2$$

Example 1.

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 - x_2 &= 3 \\ x_1 + 2x_2 &= 2 \end{aligned}$$

Graph each equation in $x_1 x_2$ plane, and check whether there exists any single intersection point.

We observe that there is no exact solution. Can we find one point that roughly satisfy in all equation?

$$\min_{x_1, x_2} \quad (1 - x_1 - x_2)^2 + (3 - x_1 + x_2)^2 + (2 - x_1 - 2x_2)^2$$

This is equivalent to

$$\min \|b - Ax\|_2^2$$

where

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix}$$

Overdetermined System: $Ax = b$, $A \in \mathbb{R}^{m \times n}$ $m > n$

- No solution $\rightarrow$ define the *residual vector*

$$r = b - Ax$$

and look for x that minimizes the length of r, i.e.,

$$\min_x \ \|r\|_2 := \|b - Ax\|_2$$

Note: The solution x^* that minimizes $\|b - Ax\|_2$ also minimizes $\|b - Ax\|_2^2$. Thus to solve overdetermined system, we always solve

$$\min_x \ \|b - Ax\|_2^2$$

$$
\begin{aligned}
f(x) &= \|b - Ax\|^2 \\
&= (Ax - b)^T (Ax - b) \\
&= x^T A^T Ax - 2b^T Ax + b^T b
\end{aligned}
$$

- This is a quadratic function
- $A^T A$ is symmetric
- $A^T A$ is positive semidefinite $x^T A^T Ax = \|A^T x\|^2 \geq 0$
- Convex function $\rightarrow$ minimum exists
- $\nabla f(x) = 2A^T Ax - (2b^T A)^T = 2A^T A - 2A^T b$
- $\nabla f(x) = 0 \rightarrow$ solve the system $A^T Ax = A^T b$
- Minimom point $x^* = (A^T A)^{-1} A^T b$

Practical Application: Linear Regression

You can view linear regression as an overdetermined linear system. Let assume there is n points

$$[x_1, y_1]^T, \dots, [x_n, y_n]^T \in \mathbb{R}^2$$

where x_i's are regressors and y_i's are respond values.

The goal of linear regression is to find the (regression) line of the form

$$y = ax + b$$

where a and b satisfy

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ \qquad \vdots \\ ax_n + b = y_n \end{cases}$$

Does this system have a unique solution for a and b?

For $n \geq 3$, the system is overdetermined. Thus to find a and b, we need to minimize the sum of the squared residuals that is

$$\min_{a,b} \sum_{i=1}^{n} (ax_i + b - y_i)^2.$$

In the vector/matrix form

$$\min_{a,b} \sum_{i=1}^{n} (ax_i + b - y_i)^2 \quad \Longleftrightarrow \quad \min_{z} \|Az - y\|_2^2 \quad \Longleftrightarrow \quad z^* = (A^T A)^{-1} A^T y$$

where

$$z = \begin{bmatrix} a \\ b \end{bmatrix}, \quad A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Regression and Prediction

Once you find the the best fitted line (by finding the optimal a and b)

$$\ell(x) = ax + b$$

into the data

$$[x_1, y_1]^T, \ \ldots, [x_n, y_n]^T \in \mathbb{R}^2.$$

The linear regression line can be used to find future events. How?
For the new regressor $\tilde{x}$, we need to evaluate $\ell(\tilde{x})$, that is

$$\ell(\tilde{x}) = a\tilde{x} + b$$

Example. A rocket motor is manufactured by bonding an igniter propellant and a sustainer propellant together inside a metal housing. The shear strength of the bond between the two types of propellant is an important quality characteristic. It is suspected that shear strength is related to the age in weeks of the batch of sustainer propellant. Twenty observations on shear strength and the age of the corresponding batch of propellant have been collected:

TABLE 2.1 Data for Example 2.1

Observation, i	Shear Strength, y_i (psi)	Age of Propellant, x_i (weeks)
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2256.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50

The scatter diagram is shown below, suggests that there is a strong statistical relationship between shear strength and propellant age, and the tentative assumption of the straight-line model $y = ax + b$ appears to be reasonable.

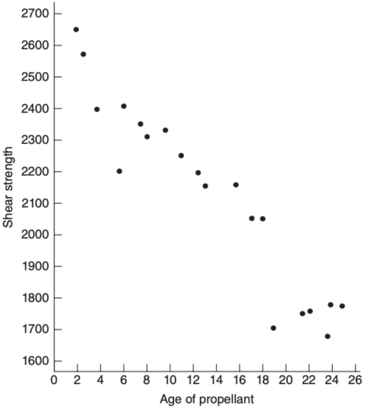

Figure 2.1 Scatter diagram of shear strength versus propellant age, Example 2.1.

Linear regression fitted line is

$$y = -37.15x + 2627.82$$

Multiple Regression Model

- More than one independent variable/ regressor
- For instance, the yield in pounds of conversion in a chemical process depends on temperature and the catalyst concentration. Then x_1 is variable for temperature and x_2 a variable for the catalyst concentration, y is the yield expressed as a linear combination of regressors x_1 and x_2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

  where $\beta_0, \beta_1, \beta_2 \in \mathbb{R}$.
- The $2d$ multiple linear regression model (has two independent variables)

$$E(x_1, x_2) = y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

  is a plane in a three dimensional space y, x_1, x_2.

Multiple Regression Model with k regressors

In this case the response y is related to k regressor or predictor variables

		Regressors			
Observation, i	Response, y	x_1	x_2	$\ldots$	x_k
1	y_1	x_{11}	x_{12}	$\ldots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\ldots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	x_{n1}	x_{n2}	$\ldots$	x_{nk}

$$y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k$$

and the model is given by

$$y = E(x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the scalar parameters, need to be found.

Let assume that we have the following set of given data

$$\left\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : i = 1, 2, \ldots, n\right\}$$

We need to find $\beta_i, i = 1, \ldots k$ such that

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_k x_{ik} = y_i, \quad i = 1, \ldots, n$$

Note that $k << n$

In vector/matrix form

$$A\beta = y$$

$$A = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ 1 & x_{31} & x_{32} & \ldots & x_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

To find β, solve

$$\min_{\beta \in \mathbb{R}^{k+1}} \|A\beta - y\|^2$$

Recall that the solution is given by

$$\beta^* = (A^T A)^{-1} A^T y$$

After you find the optimal regression line

$$E(x_1, \ldots, x_k) = \beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \ldots \beta_k^* x_k$$

For the new regressor $\tilde{x} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k] \in \mathbb{R}^k$, then the predicted respond is

$$E(\tilde{x}_1, \ldots, \tilde{x}_k) = \beta_0^* + \beta_1^* \tilde{x}_1 + \beta_2^* \tilde{x}_2 + \ldots \beta_k^* \tilde{x}_k$$