Introduction to Least Squares

Least Squares methods are crucial in chemical sciences because they provide a robust approach for data fitting, error minimization, and parameter estimation in experimental and computational studies. In fields like spectroscopy, thermodynamics, and reaction kinetics, researchers often collect noisy data, and Least Squares helps identify the best-fit model by minimizing the sum of the squared differences between observed and predicted values. This technique enables accurate characterization of molecular structures, reaction pathways, and physical properties, making it essential for reliable analysis and predictions in chemistry.

1 Linear models

Let's start by defining a class of regression models known that are linear with respect to their parameters,

$$f(\mathbf{x},\mathbf{w}) = \mathbf{w}^\top \mathbf{x} = \begin{bmatrix} w_0, & w_1, & \cdots, & w_d \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} = w_0 + \sum_{i=1}^{d} w_i x_i. \tag{1}$$

From these equation we can observe that any argument that depends on $\mathbf{x}$ is only multiply by a single w_i, therefore the name **linear models**. Polynomials expansions are also linear models,

$$f(x,\mathbf{w}) = \mathbf{w}^\top \Phi(x) = \begin{bmatrix} w_0, & w_1, & \cdots, & w_p \end{bmatrix} \begin{bmatrix} x^0 \\ x^1 \\ \vdots \\ x^p \end{bmatrix} = \sum_{i=1}^{p} w_i x^i. \tag{2}$$

Let's consider the task of fitting the Van der Waals equation by approximating it with a quadratic polynomial,

$$P(V) = c_0 + c_1 V + c_2 V^2. \tag{3}$$

For an unknown gas, we are given the following experimental data to approximate the linear parameters, $\mathbf{w}^\top = [c_0, c_1, c_2]$.

P	V
5.5	1
43.1	2
128	3

Using these data we can try to solve the following set of linear equations,

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} 5.5 \\ 43.1 \\ 128 \end{bmatrix}}_{\mathbf{y}}, \tag{4}$$

As you cam observe, matrix $\mathbf{X}$, also known as the **design matrix**, is conformed by each individual data. For this example, each data point is expanded in **polynomial features**.

As you can verify, there is no restriction into how many points we can include in this approach as the number of elements in each row are the same as the number of linear weights in $\mathbf{w}$; recall the restriction from matrix-vector multiplication.

Using Gauss Jordan elimination, let's find the values of $\mathbf{w}$.

$$\left.\begin{array}{ccc|c} 1 & 1 & 1 & 5.5 \\ 1 & 2 & 4 & 43.1 \\ 1 & 3 & 9 & 128 \end{array}\right.$$

Goal-1: Create an upper triangular matrix by manipulating each equation.
• Subtract row 2 from row 3

$$\begin{array}{ccc|c} 1 & 1 & 1 & 5.5 \\ 1 & 2 & 4 & 43.1 \\ 0 & -1 & -5 & -84.9 \end{array}$$

• Subtract row 1 from row 2, multiply row 3 by (-1)

$$\begin{array}{ccc|c} 1 & 1 & 1 & 5.5 \\ 0 & -1 & -3 & -37.6 \\ 0 & 1 & 5 & 84.9 \end{array}$$

• Sum row 2 and row 3

$$\begin{array}{ccc|c} 1 & 1 & 1 & 5.5 \\ 0 & -1 & -3 & -37.6 \\ 0 & 0 & 2 & 47.3 \end{array}$$

• We found that, $c_2 = 23.65$.

Goal-2: Create an lower triangular matrix by manipulating each equation, back-substitution.
• Multiply row 3 by (3) and sum it to row 2

$$\begin{array}{ccc|c} 1 & 1 & 1 & 5.5 \\ 0 & -1 & 0 & 33.35 \\ 0 & 0 & 1 & 23.65 \end{array}$$

• We found that, $c_1 = -33.35$.
• Subtract row (3) from row(1)

$$\begin{array}{ccc|c} 1 & 1 & 0 & -18.15 \\ 0 & 1 & 0 & -33.35 \\ 0 & 0 & 1 & 23.65 \end{array}$$

• Subtract row (2) from row(1)

$$\begin{array}{ccc|c} 1 & 0 & 0 & 15.1 \\ 0 & 1 & 0 & -33.35 \\ 0 & 0 & 1 & 23.65 \end{array}$$

• We found that, $c_0 = 15.1$.

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} 15.1 \\ -33.35 \\ 23.65 \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} 5.5 \\ 43.1 \\ 128 \end{bmatrix}}_{\mathbf{y}}, \tag{5}$$

Homework: You can do the same procedure but now for a cubic expansion, where the new term is c_3V^3. The additional data point you need is $P = 290.7$ for $V = 4$.