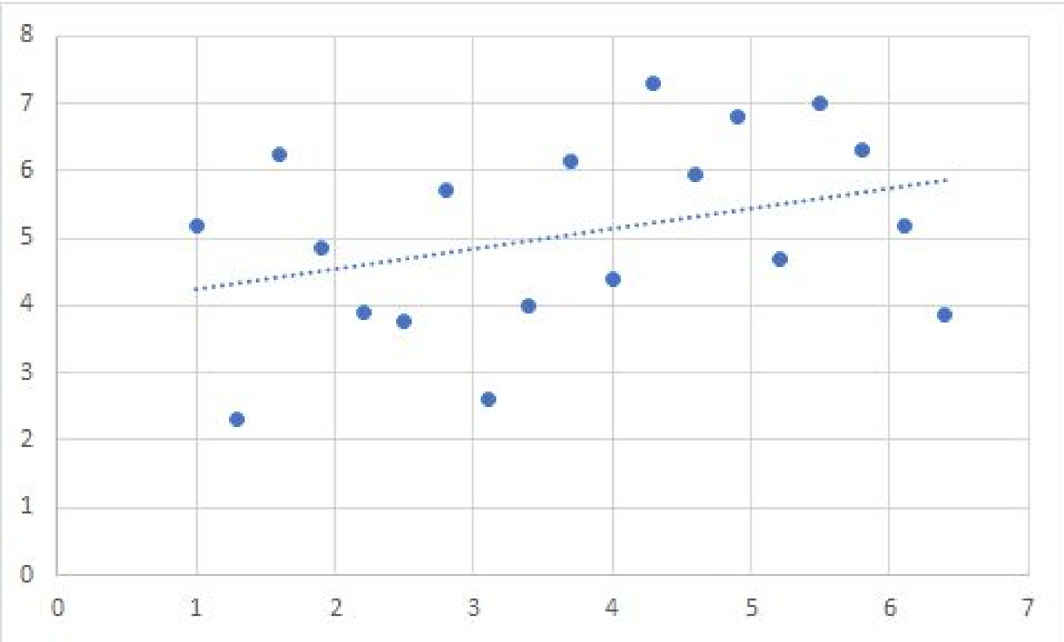
Mínimos cuadrados

¿Qué significa ajustar por mínimos cuadrados?

Seguramente, en química ya habrás tenido que calcular rectas de regresión. Dados una serie de puntos, ¿cómo encuentro la ecuación de la recta que mejor se ajusta a ellos? ¿Y si no quiero que sea una recta, sino una parábola?



Podemos ver el ajuste por mínimos cuadrados como un método para hallar los coeficientes de la recta cuya distancia a los puntos sea mínima.

Caso lineal

Supongamos que yo quiero encontrar la recta que mejor se ajusta a la nube de puntos. Toda recta es de la forma **a + bx**. Por ello, lo que tenemos que hacer es encontrar la a y la b que hacen que la suma de las distancias de los puntos a la recta sea **mínima.**

Cómo ya sabes, hallar el mínimo de una función implica el uso de derivadas. Además para facilitarnos los cálculos, vamos a elevar la distancia al cuadrado.

Dicho lo cual, empecemos: la fórmula de la distancia de un pto cualquiera a la recta, su cuadrado, y la suma de distancias al cuadrado tendrá este aspecto:

$$d_j = f(s_j) - (a + b*s_j)$$

$$d_j^2 = (f(s_j) - (a + b*s_j))^2$$

$$f(a,b) = \sum_{j=1}^{n} d_j^2 = \sum_{j=1}^{n} (f(s_j) - (a + b*s_j))^2$$

*Siendo sj y f(sj) las coordenadas de un punto de la nube de puntos

Para hallar el valor mínimo de a tendré que realizar la derivada parcial respecto a a. Lo mismo ocurre con b. Derivando parcialmente obtenemos:

$$\frac{df}{da} = 2\sum_{j=1}^{n} d_j * d_j' = 2 \sum_{j=1}^{n} (f(s_j)-(a + b*s_j))*(-1)=0$$

$$\frac{df}{db} = 2\sum_{j=1}^{n} d_j * d_j' = 2 \sum_{j=1}^{n} (f(s_j)-(a + b*s_j))*(-s_j)=0$$

*Recuerda que igualamos a cero porque quiero sacar el mínimo

Vemos que nos queda un sistema de dos ecuaciones con dos incógnitas fácilmente resoluble. Despejando el término independiente

$$\sum_{j=1}^{n} a + \sum_{j=1}^{n} b*s_j = \sum_{j=1}^{n} f(s_j) \quad \rightarrow \quad n*a + b*Sx = Sy$$

$$\sum_{j=1}^{n} a*s_j + \sum_{j=1}^{n} b*s_j^2 = \sum_{j=1}^{n} f(s_j)*s_j \quad \rightarrow \quad a*Sx + b\ Sx2 = Sxy$$

*Siendo Sx la suma de las s_j, Sx la suma de las s_j al cuadrado, Sy la suma de las $f(s_j)$ y Sxy la suma del producto de s_j por y_j.

Hemos reducido el problema a calcular 4 sumatorios (Sx, Sy, Sx2, Sxy) fácilmente expresables en un bucle.

Casos no lineales

Sin embargo, también puede ser que me interese que la nube de puntos no se ajuste a una recta, sino a un polinomio de grado más alto, o a algo que no sea un polinomio.

En cualquier caso, el procedimiento será el mismo

1. Planteo la fórmula a la que quiero ajustar la nube
2. Expreso cuánto valdría la distancia y la elevo al cuadrado
3. Derivo parcialmente la distancia respecto a cada una de las variables
4. Intento despejar para llegar a un sistema de ecuaciones con tantas incógnitas como variables.

Por ejemplo, pongamos que lo quiero ajustar a la siguiente parábola: ax^2+bx+c. De modo que la distancia será:

$$d_j=f(s_j)-(a*s_j^2 + b*s_j + c)$$
$$d_j^2=(f(s_j)-(a*s_j^2 + b*s_j + c))^2$$

Derivo la distancia en función de las tres variables a,b,c

$$\frac{df}{da} = 2\sum_{j=1}^{n} d_j * d_j' = 2 \sum_{j=1}^{n} (f(s_j)-(as_j^2 + b*s_j+c))*(-s_j^2) = 0$$

$$\frac{df}{db} = 2\sum_{j=1}^{n} d_j * d_j' = 2 \sum_{j=1}^{n} (f(s_j)-(as_j^2 + b*s_j+c))*(-s_j) = 0$$

$$\frac{df}{dc} = 2\sum_{j=1}^{n} d_j * d_j' = 2 \sum_{j=1}^{n} (f(s_j)-(as_j^2 + b*s_j+c))*(-1) = 0$$

Igual que antes, vamos a adaptar estas ecuaciones

$$\mathbf{a} \sum_{j=1}^{n} \mathbf{s_j^4} + \mathbf{b} \sum_{j=1}^{n} \mathbf{s_j^3} + \mathbf{c} \sum_{j=1}^{n} \mathbf{s_j^2} = \sum_{j=1}^{n} \mathbf{f(s_j)*s_j^2}$$

$$\mathbf{a} \sum_{j=1}^{n} \mathbf{s_j^3} + \mathbf{b} \sum_{j=1}^{n} \mathbf{s_j^2} + \mathbf{c} \sum_{j=1}^{n} \mathbf{s_j} = \sum_{j=1}^{n} \mathbf{f(s_j)*s_j}$$

$$\mathbf{a} \sum_{j=1}^{n} \mathbf{s_j^2} + \mathbf{b} \sum_{j=1}^{n} \mathbf{s_j} + \mathbf{n*c} = \sum_{j=1}^{n} \mathbf{f(s_j)}$$

De nuevo llegamos a un sistema de ecuaciones, esta vez, 3 ecuaciones con 3 incógnitas. Vemos que ahora es más complicado resolverlo directamente; y según aumentemos el grado más difícil será. Por ello, es importante que, a la hora de resolverlo sepamos utilizar Cramer

Método de Cramer

Para empezar, debemos crearnos una matriz con los coeficientes del sistema:

$$A = \begin{pmatrix} Sx4 & Sx3 & Sx2 \\ Sx3 & Sx2 & Sx \\ Sx2 & Sx & n \end{pmatrix}$$

Bien, Cramer nos dice que la a será igual al determinante de la matriz A **PERO** sustituyendo la columna de coeficientes de la a por la columna de términos independientes. Todo esto dividido por el determinante de A. Para la b, sustituiré la columna de las b y para la c lo mismo, de forma que el resultado será:

$$a = \frac{\begin{vmatrix} Syx2 & Sx3 & Sx2 \\ Syx & Sx2 & Sx \\ Sy & Sx & n \end{vmatrix}}{|A|} \qquad b = \frac{\begin{vmatrix} Sx4 & Syx2 & Sx2 \\ Sx3 & Syx & Sx \\ Sx2 & Sy & n \end{vmatrix}}{|A|} \qquad c = \frac{\begin{vmatrix} Sx4 & Sx3 & Syx2 \\ Sx3 & Sx2 & Syx \\ Sx2 & Sx & Sy \end{vmatrix}}{|A|}$$

OJO: El proceso será el mismo para grados más altos, es decir, sistemas con más incógnitas.