

Universidad Nacional de Colombia

FACULTAD DE CIENCIAS

MÉTODO DE LOS MÍNIMOS
CUADRADOS

Tarea 2 Análisis Numérico

Autor:

Shara Gallego Grisales, Diego Carvajal

Enero 2025

Índice

1. Introducción	2
2. Teoría	3
2.1. ¿Qué es el Método de Mínimos Cuadrados?	3
3. Perspectiva Geométrica	3
4. Perspectiva Algebraica	5
5. Perspectiva de Data Science	5
6. Conclusiones	7
7. Bibliografía	8

1. Introducción

El método de mínimos cuadrados es una herramienta fundamental en el análisis numérico y la estadística, utilizada para encontrar la mejor aproximación a un conjunto de datos cuando no es posible una solución exacta. Su aplicación abarca desde la resolución de sistemas sobredeterminados hasta la regresión de datos en ciencia de datos y aprendizaje automático.

En este documento, se exploran diversas perspectivas del método: su formulación matemática, su interpretación geométrica como una proyección sobre el espacio columna, y su aplicación en ciencia de datos para realizar regresiones.

A través de estas perspectivas, se busca proporcionar una comprensión integral de este método y su relevancia en el análisis de datos y la optimización.

2. Teoría

2.1. ¿Qué es el Método de Mínimos Cuadrados?

Al enfrentarnos al siguiente problema $Ax = b$ donde $A \in \mathbb{R}^{m \times n}$ y conocemos el vector de datos b , el método de mínimos cuadrados ofrece una forma de minimización del vector residual $r = b - Ax$ minimizando la norma al cuadrado del vector residual $\langle r, r \rangle$.

Esta forma de optimizar el error es atractiva por varias razones; una de ellas es que la función de $\|.\|^2$ es una función suave.

Entonces es fácil ver que dados los datos este método de una manera muy eficaz nos da la "mejor aproximación" posible para todos los puntos a la cual llamaremos modelo

3. Perspectiva Geométrica

Una manera interesante de ver el método de mínimos cuadrados es notar que la solución de este problema se deja ver como la proyección de b en el espacio columna de A .

Ahora cuando el sistema $Ax = b$ (donde b es el vector dado por los datos) no tiene una solución exacta, es decir que $b \notin \text{col}(A)$ y buscamos su mejor aproximación, utilizamos el método de mínimos cuadrados para minimizar el vector residual $r = b - Ax$.

Ahora si tomamos el vector $\bar{b} \in \text{col}(A)$ como el vector más cercano a b en el subespacio $\text{col}(A)$ tendríamos que \bar{b} es una buena aproximación que

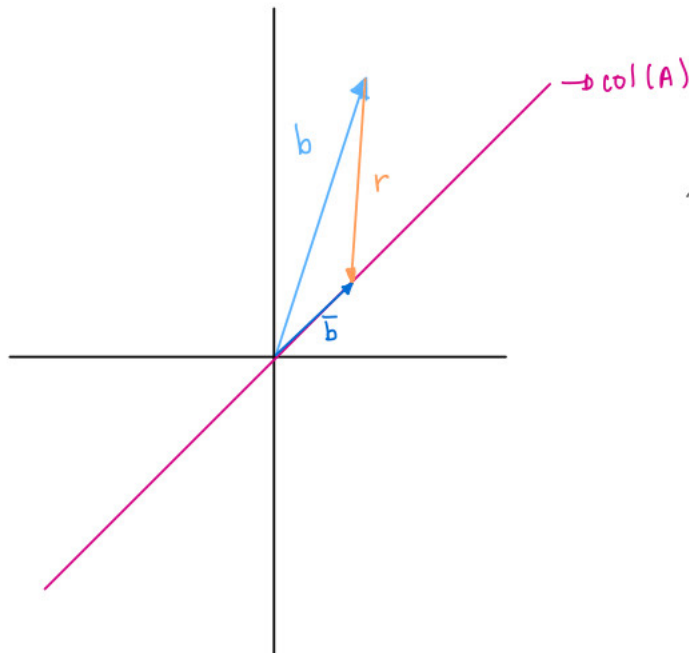
soluciona nuestro sistema, este vector \bar{b} podemos mostrar que se deja ver como la proyección de b en el subespacio $col(A)$ tal que cumple lo siguiente

$$\min_{x \in col(A)} \|b - x\|$$

Note que si $\bar{b} = Ax$ por lo tanto $r = b - \bar{b}$ luego por definición de la proyección se sigue que r es ortogonal a $col(A)$ es decir que $A^T r = 0$

Asumimos la ortogonalidad de r pues si r no fuera ortogonal a \bar{b} esto significaría que existe una mejor aproximación en $col(A)$ para b por lo tanto no cumpliría la condición de minimización y la proyección se deduciría de este hecho.

Así tendríamos que $\bar{b} = Ax$, por definición tendríamos que $r = b - \bar{b}$ es decir que $b = r + \bar{b}$, si aplicáramos la propiedad del triángulo esto se vería de la siguiente manera.



4. Perspectiva Algebraica

La ecuación normal del método de mínimos cuadrados se obtiene de minimizar la suma de los vectores residuo al cuadrado.

Supongamos que $Ax = \bar{b}$ define el modelo para nuestra regresión lineal.

Ahora nuestro objetivo es minimizar el cuadrado de la norma del residual, que llamaremos $S(x)$. Así:

$$\begin{aligned} S(x) &= \langle r, r \rangle \\ &= (b - \bar{b})^T (b - \bar{b}) \\ &= (b - Ax)^T (b - Ax) \\ &= b^T b - b^T Ax - x^T A^T b + x^T A^T Ax \\ &= b^T b - 2x^T A^T b + x^T A^T Ax \end{aligned}$$

Ahora minimicemos $S(x)$

$$\begin{aligned} \frac{\partial S(x)}{\partial x} &= -2A^T b + 2A^T Ax = 0 \\ &\iff A^T Ax = A^T b \end{aligned}$$

Ahora suponiendo que $A^T A$ es invertible tendríamos que

$$x = (A^T A)^{-1} A^T b \quad (1)$$

A la ecuación (1) la llamaremos la ecuación normal del método de mínimos cuadrados

5. Perspectiva de Data Science

Uno de los usos más importantes de mínimos cuadrados es hacer regresiones. Dado un conjunto de datos $(x_1, y_i)_{i=1}^n \in \mathbb{R}^2$, podemos tratar de modelar funciones que describan aproximadamente el comportamiento de estos datos. Ya vimos una estrategia en el curso de hacer interpolación con polinomios o splines, encontrando funciones $f(x)$ que cumplan que $f(x_i) = y_i \quad \forall i$ y sean

bien comportadas en ciertos sentidos. A pesar de que el hecho de que interpolen pueda ser deseable en algunos contextos, esta idea puede llegar a tener problemas. Por ejemplo, al usar interpolación de Lagrange o de Hermite, si se tiene una cantidad considerable de datos, por la naturaleza oscilante de los polinomios de grado alto, se puede llegar a tener problemas numéricos y funciones de predicción feas. Además, se suelen tener que definir en un intervalo muy cercano alrededor de cada dato, pues su comportamiento lejos de ellos puede ser impredecible, y al tratar de interpolar funciones acotadas, esta interpolación solo es buena en un intervalo, pues al usar polinomios, que crecen o decrecen tanto como se quiera alejándose del origen, eventualmente sus valores crecen o decrecen demasiado como para seguir siendo aproximaciones válidas.

Sin embargo, con una regresión, podemos encontrar más bien funciones que se acerquen a estos puntos dados y sean las "mejores" que lo hagan en el sentido de que minimicen la suma de los cuadrados de las distancias entre los puntos y la función evaluada en los puntos, es decir, la suma de los cuadrados de los residuales. Veamos como hacerlo.

Digamos que queremos modelar los datos con una función lineal $f(x) = \beta_0 + \beta_1 x$, donde β_0, β_1 son parámetros a determinar. Queremos tomarlos tales que se acerquen lo más posible a los datos. Lo mejor posible es que los interpolen, es decir, $f(x_i) = y_i \quad \forall i$, que podemos ver en forma matricial $Ax = b$ como:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

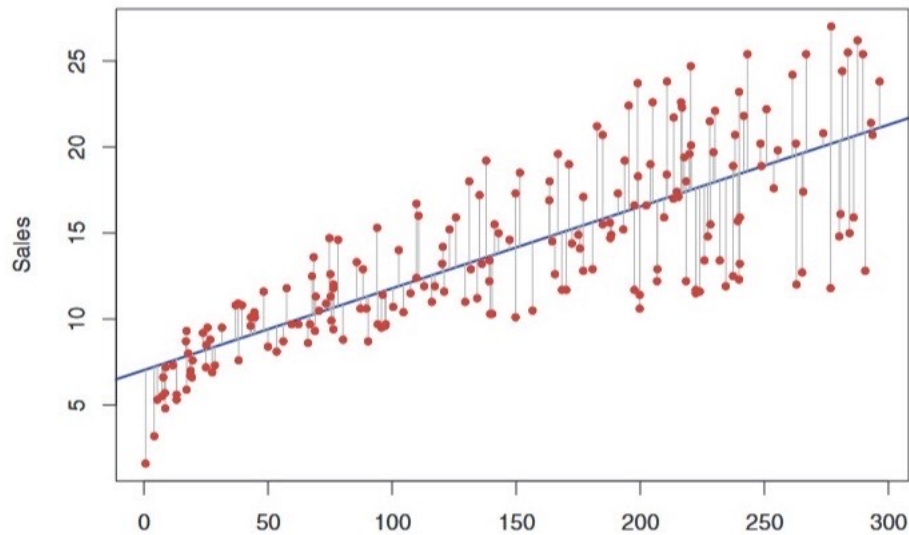
Sin embargo, esto casi nunca se puede alcanzar, pues tenemos muchísimos más ecuaciones que incógnitas. Aquí entra entonces mínimos cuadrados. Resolvemos las ecuaciones normales $A^T Ax = A^T b$ para encontrar el vector $\hat{x} = [\hat{\beta}_0, \hat{\beta}_1]^T$ que minimice $\|b - Ax\|$. Con esto, obtenemos la regresión $f(x) = \hat{\beta}_0 + \hat{\beta}_1 x$ que más se acerca a los datos dados.

Podemos generalizar esto a funciones más complicadas que una función lineal, siempre y cuando sean *lineales* en los parámetros: $f(x) = \sum_{j=0}^m \beta_j g_j(x) \quad (m <$

$n - 1$), resolviendo el siguiente sistema en mínimos cuadrados:

$$\begin{bmatrix} g_0(x_1) & g_1(x_1) & \cdots & g_m(x_1) \\ g_0(x_2) & g_1(x_2) & \cdots & g_m(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ g_0(x_n) & g_1(x_n) & \cdots & g_m(x_n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Esto permite mucha más libertad que una simple interpolación, pudiendo usar más clases de funciones para aproximar los datos



6. Conclusiones

- El método de mínimos cuadrados es equivalente a proyectar el vector de los datos en el espacio columna de la matriz asociada a nuestro sistema lineal
- La ecuación normal es la derivada de la suma de los vectores residuos
- Permite hacer regresiones para acomodar nubes de datos dados a varios tipos de funciones, tratando de minimizar la distancia total entre los puntos dados y la predicción

7. Bibliografía

Referencias

- [1] D. Bindel, *Lecture 10: Numerical Analysis*, Cornell University, 2012. Disponible en: <https://www.cs.cornell.edu/~bindel/class/cs3220-s12/notes/lec10.pdf>.
- [2] R. L. Burden, J. D. Faires, *Numerical Analysis, 9th Edition*, Cengage Learning, 2010. Disponible en: https://faculty.ksu.edu.sa/sites/default/files/numerical_analysis_9th.pdf.
- [3] Wikipedia contributors, *Linear Least Squares*, Wikipedia, The Free Encyclopedia. Disponible en: https://en.wikipedia.org/wiki/Linear_least_squares.
- [4] Wikipedia contributors, *Vandermonde Matrix*, Wikipedia, The Free Encyclopedia. Disponible en: https://en.wikipedia.org/wiki/Vandermonde_matrix.