Análisis Numérico

Tarea 2: Mínimos cuadrados y aplicaciones

Teoría de los mínimos cuadrados

El método de Mínimos Cuadrados, introducido por primera vez en el siglo XIX, se atribuye a los matemáticos Adrien-Marie Legendre y Carl Friedrich Gauss quienes lo desarrollaron en contextos distintos: uno en problemas de ajuste de datos y el otro en predicción astronómica.

Esta es una técnica fundamental en matemáticas y estadística que se utiliza para encontrar la función que mejor se aproxime a un conjunto de datos observados. Su fundamento matemático se basa en minimizar la suma de los cuadrados de las diferencias entre los valores observados ($\mathbf{y}$) y los estimados ($\hat{\mathbf{y}} = A\hat{\beta}$), conocida como la norma cuadrática del residuo $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$. En términos de álgebra lineal, se trata de aproximar soluciones a sistemas lineales $\mathbf{y} = A\beta$ que son sobredeterminados, es decir, con más ecuaciones que incógnitas.

Si se tiene un sistema $\mathbf{y} = A\beta$ sin solución exacta, lo mejor que se puede hacer es encontrar una solución aproximada $\hat{\beta}$ tal que el vector proyectado $\hat{\mathbf{y}} = A\hat{\beta}$ sea lo más cercano posible a $\mathbf{y}$. Esto equivale a minimizar la distancia perpendicular entre el vector original $\mathbf{y}$ y el subespacio $Col(A)$. Y es así como el método encuentra la proyección ortogonal de $\mathbf{y}$ sobre $Col(A)$, lo cual garantiza que el modelo ajustado minimice la suma de los cuadrados de las componentes del residuo $\mathbf{r}$. Este procedimiento conduce a las ecuaciones normales que permiten determinar los coeficientes que mejor ajustan los datos.

A lo largo de los años, por su simplicidad y efectividad, el método de los mínimos cuadrados es sumamente importante en diversas áreas de la ciencia y la ingeniería.

Perspectiva geométrica

El método de los mínimos cuadrados se basa en el concepto de proyección ortogonal en espacios vectoriales donde se minimiza la distancia perpendicular entre los puntos de datos y el modelo ajustado.

En álgebra lineal, se conoce la proyección ortogonal de un vector sobre un subespacio como el vector más cercano al original que pertenece dicho subespacio. Si un vector $\mathbf{y}$ está en un espacio vectorial $\mathbb{R}^n$ y se desea proyectarlo sobre un subespacio S generado por algún conjunto de vectores, entonces el resultado de esta proyección es un vector $\hat{\mathbf{y}} \in S$ tal que el residuo $\mathbf{r} = \mathbf{r} - \hat{\mathbf{y}}$ es ortogonal a S. Esto asegurará que la proyección minimiza la distancia $||\mathbf{r}||$; en otras palabras, el residuo r tiene la menor magnitud posible. Geométricamente, esto es encontrar el punto más cercano en el subespacio al vector original.

En el caso de la regresión lineal se desea ajustar una línea de la forma $y = \beta_0 + \beta_1 x$ y para ello se puede expresar el problema en términos de álgebra lineal tal y como se explicó anteriormente:

Dado un conjunto de n puntos (x_i, y_i) donde los valores observados y_i se agrupan en un vector $\mathbf{y}$ y la matriz A tiene una columna de unos y una columna con los valores x_i.

El objetivo es encontrar el vector $\hat{\beta} = [\beta_0, \beta_0]^\top$ que minimice la norma del residuo:

$$\mathbf{r} = \mathbf{y} - A\hat{\beta}$$

Esto se traduce en proyectar el vector $\mathbf{y}$ sobre el subespacio generado por las columnas de A. Y así, el resultado $\hat{\mathbf{y}}$ es la proyección ortogonal de $\mathbf{y}$ y el residuo $\mathbf{r}$ es el componente ortogonal a dicho subespacio. En otras palabras, la proyección es ortogonal porque los valores predichos y los valores reales no están correlacionados. Esto se ilustra en la Figura 1, que representa el caso de dos variables independientes (vectores $\mathbf{x_1}$ y $\mathbf{x_2}$) y el vector de datos ($\mathbf{y}$), y muestra que el vector de error ($\mathbf{y_1} - \hat{\mathbf{y}}$) es ortogonal a la estimación del mínimo cuadrado ($\hat{\mathbf{y}}$) que se encuentra en el subespacio definido por las dos variables independientes.

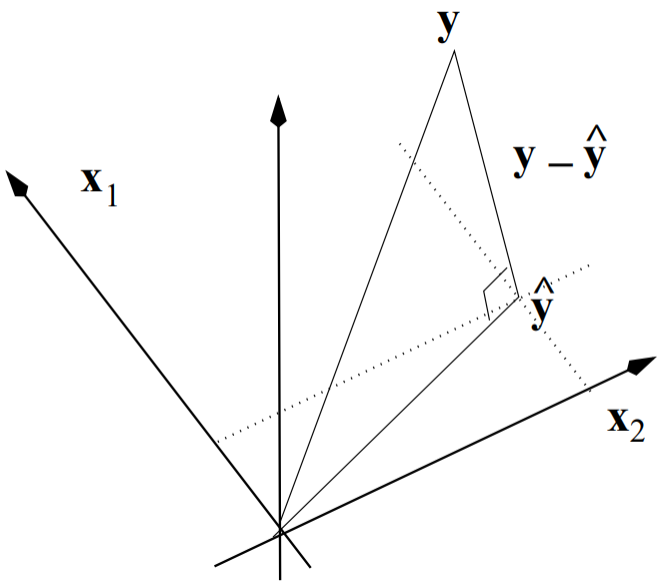

Figura 1: *La estimación de los mínimos cuadrados de los datos es la proyección ortogonal del vector de datos sobre el subespacio de la variable independiente.*

En el caso polinómico, el objetivo es ajustar un polinomio de grado n a un conjunto de datos (x_i, y_i). El polinomio tiene la forma:

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

Cada función $1, x, x^2, \ldots, x^n$ genera un vector en un espacio de dimensión igual al número de observaciones m. Al evaluar dichas funciones en los puntos x_i se obtiene una matriz A cuya columna j-ésima corresponde a la función x^{j-1}. El problema de los mínimos cuadrados consistiría en este caso en encontrar los coeficientes $a = [a_0, a_1, \ldots, a_n]^\top$ que minimicen la distancia entre $\mathbf{y}$ y el subespacio generado por las columnas de A. Análogamente al caso lineal, la solución es una proyección ortogonal y el residuo $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ es ortogonal al subespacio.

En esencia, los mínimos cuadrados pretende minimizar la distancia perpendicular entre un vector de datos $\mathbf{y}$ y el modelo ajustado $\hat{\mathbf{y}}$. Matemáticamente, implica resolver:

$$\min_{\beta} \|\mathbf{y} - A\hat{\beta}\|^2.$$

donde el término $||\mathbf{y} - A\hat{\beta}||^2$ representa la suma de los cuadrados de los residuos que son las distancias perpendiculares desde los puntos observados al modelo. La solución se encontraría al resolver las ecuaciones normales $A^\top A\hat{\beta} = A^\top \mathbf{y}$ lo que garantiza que $\hat{\mathbf{y}}$ es la proyección ortogonal de $\mathbf{y}$, cuya derivación y justificación se hará más adelante en la sección **Perspectiva algebraica.**

A continuación, en la Figura 2, se muestra concretamente cómo funciona visualmente el método.

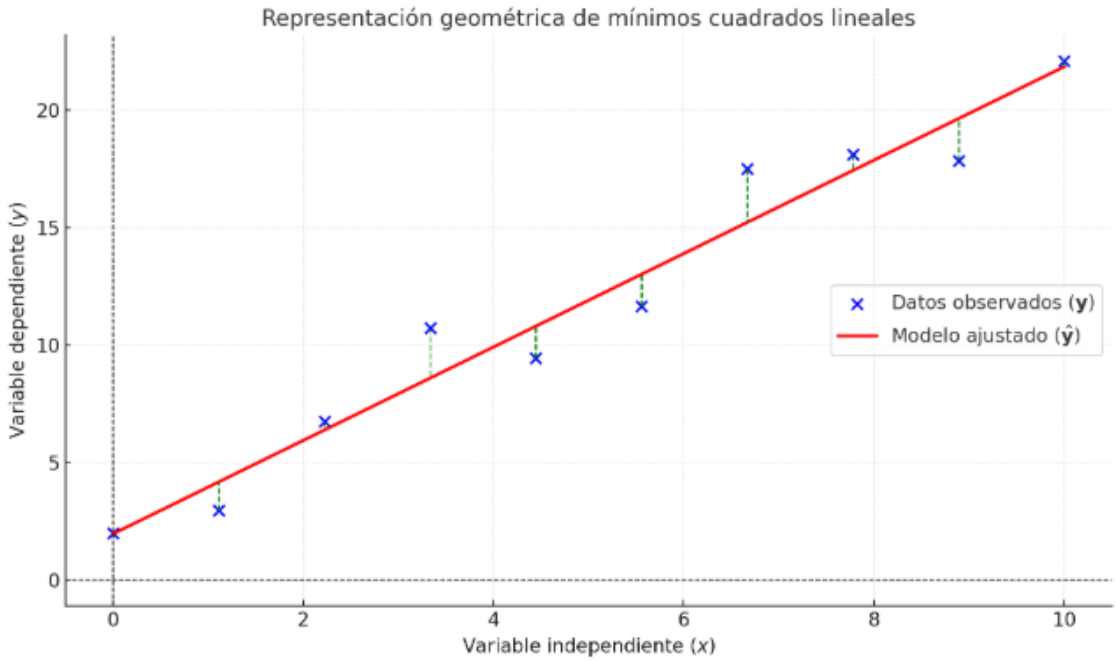

Figura 2: Mínimos cuadrados.

- **Puntos azules.** Representan los datos observados ($\mathbf{y}$)
- **Línea roja.** Modelo ajustado ($\hat{\mathbf{y}}$) que es la proyección ortogonal de los datos sobre el subespacio definido por las columnas de la matriz A.
- **Líneas verdes.** Representan los residuos $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$, es decir, las diferencias entre los valores observados y los ajustados.

Perspectiva algebraica

Deducción general de las ecuaciones normales

Inicialmente, veamos la deducción de las ecuaciones normales para cualquier sistema lineal de ecuaciones sobredeterminado donde el modelo ajustado tiene la forma $\mathbf{y} = A\beta$. Como vimos en la sección anterior, en los mínimos cuadrados se pretende resolver:

$$\min_{\beta} \|\mathbf{y} - A\hat{\beta}\|^2.$$

donde:

- $\mathbf{y}$ es el vector de datos observados.
- A es la matriz de diseño, $A \in \mathbb{R}^{mxn}$ cuyas funciones base sería, por ejemplo, $1, x, x^2, \ldots,$ etc.
- $A\hat{\beta}$ es el modelo ajustado que representa la proyección de $\mathbf{y}$ sobre $Col(A)$ y $\hat{\beta} \in \mathbb{R}^n$ es un vector de parámetros a determinar.
- $\mathbf{r} = \mathbf{y} - A\hat{\beta}$ es el residuo que se quiere minimizar.

Si se expande $\|\mathbf{y} - A\hat{\beta}\|^2$ resulta:

$$(\mathbf{y} - A\hat{\beta})^\top(\mathbf{y} - A\hat{\beta}) = \mathbf{r}^\top\mathbf{r}$$

Y al desarrollar se sigue:

$$\mathbf{y}^\top\mathbf{y} - 2\hat{\beta}^\top A^\top\mathbf{y} + \hat{\beta}^\top A^\top\hat{\beta}$$

Para encontrar el mínimo, derivamos con respecto a $\hat{\beta}$ y se iguala a cero:

$$\frac{\partial}{\partial\hat{\beta}}[\mathbf{y}^\top\mathbf{y} - 2\hat{\beta}^\top A^\top\mathbf{y} + \hat{\beta}^\top A^\top\hat{\beta}] = 0$$

$$-2A^\top\mathbf{y} + 2A^\top A\hat{\beta} = 0$$
$$2A^\top A\hat{\beta} = 2A^\top\mathbf{y}$$
$$\therefore A^\top A\hat{\beta} = A^\top\mathbf{y}$$

Estas son las llamadas **ecuaciones normales** donde su solución (o soluciones) $\hat{\beta}$ proporciona los coeficientes del modelo ajustado.

La deducción anterior asegura que $\hat{\mathbf{y}} = A\hat{\beta}$ es la proyección ortogonal de $\mathbf{y}$ sobre el subespacio columna de A, $Col(A)$.

$$\mathbf{r} = \mathbf{y} - A\hat{\beta} \text{ es ortogonal a } Col(A)$$

Algebraicamente, esto significa que: $A^\top\mathbf{r} = A^\top(\mathbf{y} - A\hat{\beta}) = 0$.

Note que hay una conexión evidente con la **perspectiva geométrica** pues el vector residuo $\mathbf{r}$ es perpendicular al subespacio columna de A y la solución $\hat{\beta}$ minimiza la distancia entre $\mathbf{y}$ y $A\hat{\beta}$.

Caso particular: ajuste lineal por mínimos cuadrados.

Ahora, veamos las ecuaciones normales en el caso de que el modelo ajustado sea una línea recta $y = a_1x + a_0$. Para este caso, las ecuaciones normales se obtienen para determinar los coeficientes a_0 (intersección) y a_1 (pendiente). Los datos observados son los puntos (x_i, y_i) donde:

- x_i es el valor de la variable independiente (entradas de la primera columna de la matriz A).

- y_i es el valor de la variable dependiente (los valores del vector $\mathbf{y}$).

Es así como las matriz A y el vector β se construyen de la siguiente manera:

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}, \beta = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \mathbf{y} = [y_1, y_2, \ldots, y_m]^\top$$

Ahora bien, para ajustar la mejor línea a una colección de datos, como ya vimos en la deducción general, consiste en minimizar el error total $E(\beta)$ que sería:

$$E(\beta) = \|\mathbf{y} - A\beta\|^2 = \sum_{i=1}^{m}(y_i - (a_1 x_i + a_0))^2$$

Donde m es el número total de datos observados. Para minimizar el término y encontrar los coeficientes a_1 y a_0 correspondientes se necesita que:

$$\frac{\partial E}{\partial a_0} = 0$$

y

$$\frac{\partial E}{\partial a_1} = 0$$

esto es,

$$0 = \frac{\partial E}{\partial a_0}\sum_{i=1}^{m}(y_i - (a_1 x_i + a_0))^2 = 2\sum_{i=1}^{m}(y_i - a_1 x_i - a_0)(-1) \tag{1}$$

y

$$0 = \frac{\partial E}{\partial a_1}\sum_{i=1}^{m}(y_i - (a_1 x_i + a_0))^2 = 2\sum_{i=1}^{m}(y_i - a_1 x_i - a_0)(-x_i) \tag{2}$$

De (1) se sigue que:

$$0 = 2\sum_{i=1}^{m}(y_i - a_1 x_i - a_0)(-1)$$

$$0 = \sum_{i=1}^{m}(y_i - a_1 x_i - a_0)(-1)$$

$$0 = -\sum_{i=1}^{m} y_i + \sum_{i=1}^{m} a_1 x_i + \sum_{i=1}^{m} a_0$$

$$\sum_{i=1}^{m} y_i = a_1 \sum_{i=1}^{m} x_i + a_0 m \tag{3}$$

Y de (2):

$$0 = 2\sum_{i=1}^{m}(y_i - a_1 x_i - a_0)(-x_i)$$

$$0 = \sum_{i=1}^{m}(y_i - a_1 x_i - a_0)(-x_i)$$

$$0 = -\sum_{i=1}^{m} x_i y_i + \sum_{i=1}^{m} a_1 x_i^2 + \sum_{i=1}^{m} a_0 x_i$$

$$\sum_{i=1}^{m} x_i y_i = a_1 \sum_{i=1}^{m} x_i^2 + a_0 \sum_{i=1}^{m} x_i \tag{4}$$

Las ecuaciones (3) y (4) serían entonces las **ecuaciones normales** y de allí se despejan a_0 y a_1 para hallar la solución al sistema de ecuaciones:

$$a_0 = \frac{\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} x_i y_i \sum_{i=1}^{m} x_i}{m(\sum_{i=1}^{m} x_i^2) - (\sum_{i=1}^{m} x_i)^2}$$

y

$$a_1 = \frac{m\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m(\sum_{i=1}^{m} x_i^2) - (\sum_{i=1}^{m} x_i)^2}$$

Caso particular: ajuste polinómico por mínimos cuadrados.

Finalmente, veamos las ecuaciones normales en el caso de que el modelo ajustado sea un polinomio de grado n :

$$P_n(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1} + a_n x^n$$

En este caso, la matriz A contiene las potencias de las variables independientes x_i que corresponden a los términos del polinomio. Si el polinomio tiene grado n y hay m puntos observados $(x_1, y_1), \ldots (x_m, y_m)$ la matriz A tiene m filas y $n+1$ columnas:

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix},$$

Cada fila representaría un punto observado x_i y cada columna representa un término del polinomio. Por otra parte, el vector β contiene los coeficientes del polinomio que se quiere estimar:

$$\beta = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Y el vector $\mathbf{y}$ contiene los valores observados de la variable dependiente y_i:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Nuevamente, el objetivo es minimizar el error total $E(\beta) = \|\mathbf{y} - A\beta\|^2$, que en este contexto, equivalentemente, sería:

$$E = \sum_{i=1}^{m} (y_i - P_n(x_i))^2$$

Expandamos la ecuación para derivar parcialmente e igualar a cero.

$$\sum_{i=1}^{m} (y_i - P_n(x_i))^2 = \sum_{i=1}^{m} (y_i)^2 - 2\sum_{i=1}^{m} P_n(x_i)y_i + \sum_{i=1}^{m} (Pn(x_i))^2$$

$$= \sum_{i=1}^{m} (y_i)^2 - 2\sum_{i=1}^{m} (\sum_{j=0}^{n} a_j x_i^j) y_i + \sum_{i=1}^{m} (\sum_{j=0}^{n} (a_j x_i^j)^2$$

$$= \sum_{i=1}^{m} (y_i)^2 - 2\sum_{j=0}^{n} a_j (\sum_{i=1}^{m} y_i x_i^j) + \sum_{j=0}^{n}\sum_{j=0}^{n} a_j a_k (\sum_{i=1}^{m} (x_i)^{j+k})$$

Así como en el caso lineal, para que E se minimice es necesario que $\frac{\partial E}{\partial a_j} = 0$ para cada $j = 0, 1, \ldots, n$. Por lo cual, para cada j:

$$0 = \frac{\partial E}{\partial a_j} = -2\sum_{i=1}^{m} y_i (x_i)^j + 2\sum_{k=0}^{n} a_k \sum_{i=1}^{m} (x_i)^{j+k} \tag{5}$$

De (5) se generan $n+1$ ecuaciones normales en las $n+1$ incógnitas a_j. Estas son:

$$a_0 \sum_{i=1}^{m} x_i^0 + a_1 \sum_{i=1}^{m} x_i^1 + a_2 \sum_{i=1}^{m} x_i^2 + \cdots + a_n \sum_{i=1}^{m} x_i^n = \sum_{i=1}^{m} y_i x_i^0$$

$$a_0 \sum_{i=1}^{m} x_i^1 + a_1 \sum_{i=1}^{m} x_i^2 + a_2 \sum_{i=1}^{m} x_i^3 + \cdots + a_n \sum_{i=1}^{m} x_i^{n+1} = \sum_{i=1}^{m} y_i x_i^1$$

$$\vdots$$

$$a_0 \sum_{i=1}^{m} x_i^n + a_1 \sum_{i=1}^{m} x_i^{n+1} + a_2 \sum_{i=1}^{m} x_i^{n+2} + \cdots + a_n \sum_{i=1}^{m} x_i^{2n} = \sum_{i=1}^{m} y_i x_i^n$$

Esto corresponde al desarrollo explícito de $A^\top A$ y $A^\top \mathbf{y}$. Estas ecuaciones normales tienen solución única siempre y cuando las x_i sean distintas.

Perspectiva desde la ciencia de datos

Si se tiene un conjunto de datos y se quiere establecer una relación entre ellos o ser capaces de predecir/estimar futuros valores, se puede aproximar una relación lineal utilizando una técnica conocida como regresión lineal.

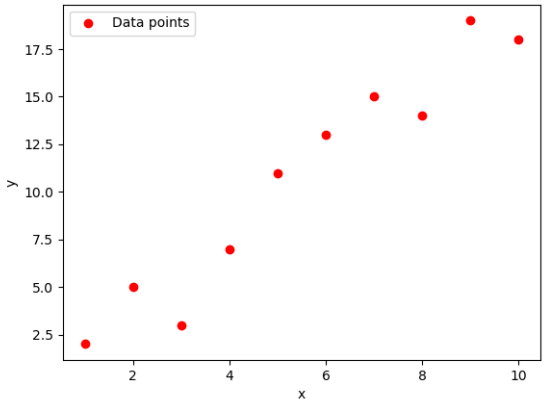

Figura 3: *Conjunto de puntos a predecir/estimar*

La regresión lineal entonces busca una relación lineal $y = mx + b$ tal que los parametros m y b sean los idoneos para el conjunto de datos. Si suponemos que los valores observdos son de la forma y_i y los valores estimados de la forma $\hat{y}_i = mx_i + b$, entonces para hallar m y b se utiliza el método de minimos cuadrados, minimizando la suma de los cuadrados de la diferencia de los valores observados y los valores estimados, esto es:

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Una vez encontrados los paramétros tenemos una recta que minimiza los datos conocidos que podemos usar para estimar nueva información

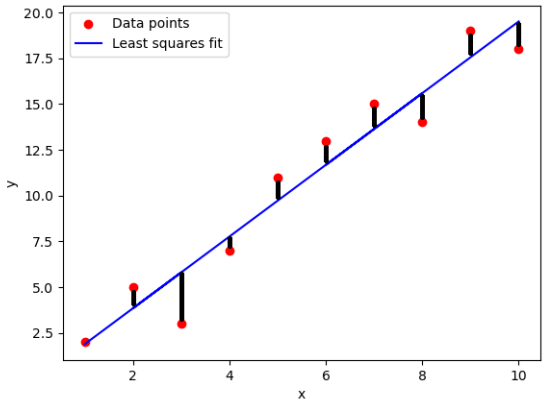

Figura 4: *Ajuste lineal*

Los minimos cuadrados son usados en la ciencia de datos moderna no solo para la regresión lineal sino también para optimizar la función de perdida en métodos de machine learning como Ridge y Lasso.

Descripción del problema

Variable Name	Role	Type	Description	Units
No	ID	Integer	Identification number	None
X1 transaction date	Feature	Continuous	Transaction date (e.g., 2013.250 = 2013 March, 2013.500 = 2013 June)	None
X2 house age	Feature	Continuous	Age of the house	Years
X3 distance to the nearest MRT station	Feature	Continuous	Distance to the nearest Mass Rapid Transit station	Meters
X4 number of convenience stores	Feature	Integer	Number of convenience stores within walking distance	Count
X5 latitude	Feature	Continuous	Geographic latitude coordinate	Degrees
X6 longitude	Feature	Continuous	Geographic longitude coordinate	Degrees
Y house price of unit area	Target	Continuous	House price per unit area	10000 New Taiwan Dollar/Ping (303.07 USD/3.3 Squared Meters)

Figura 5: *Yeh, I. (2018). Real Estate Valuation [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5J30W*

Ser capaces de predecir los precios de los bienes inmobiliarios es de altísima importancia para múltiples intereses, sean compradores, vendedores o inversores. La capacidad de predecir con exactitud los precios beneficia a la estabilidad económica, la planificación financiera individual y el desarrollo estratégico en entornos urbanos, por ello, se entrena un modelo de machine learning usando el dataset Real Estate Valuation citado anteriormente (real-state-valuation.xlsx) para ser capaces de predecir el precio de un bien inmueble (el precio por unidad de área) usando los parámetros que se pueden observar en la Figura 5. En problemas de regresión como estos se puede definir la función de pérdida usando mínimos cuadrados, a continuación veremos una implementación de dicho modelo usando el dataset ya mencionado.

Análisis de resultados

A continuación dispondremos las gráficas de la visualización de resultados para basarnos en ellas y realizar el análisis de los mismos.

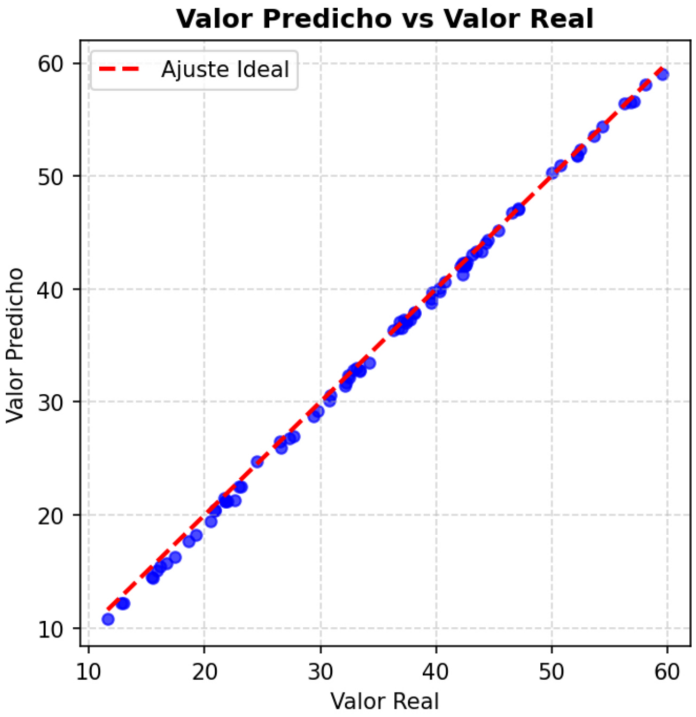

Figura 6

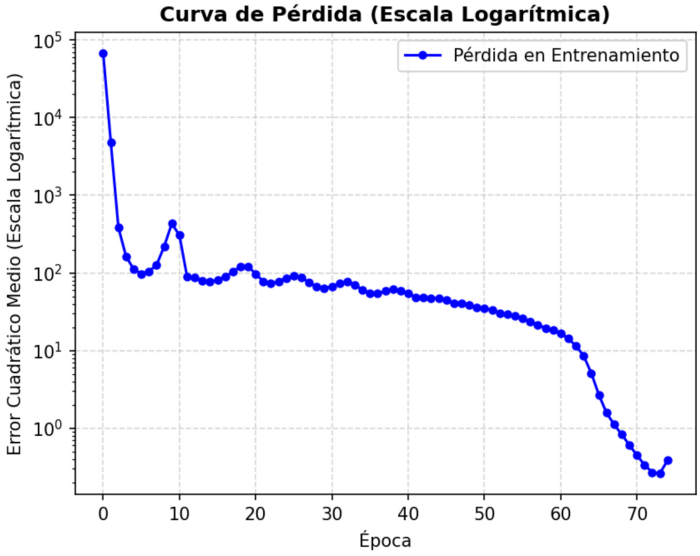

Figura 7

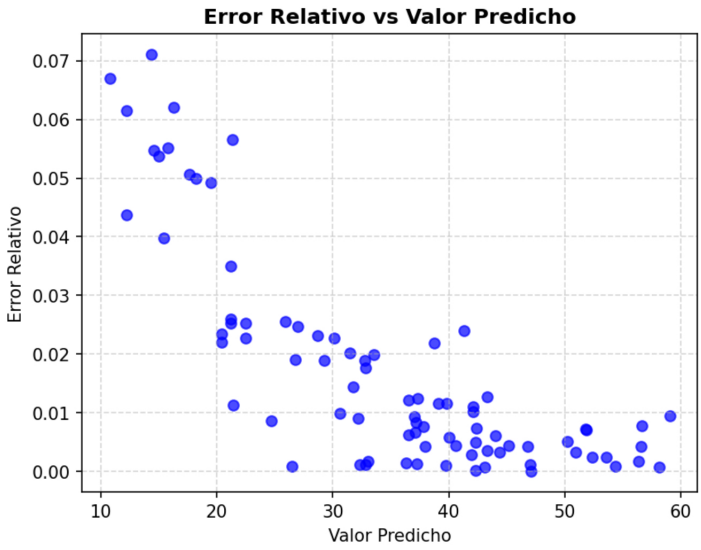

Figura 8

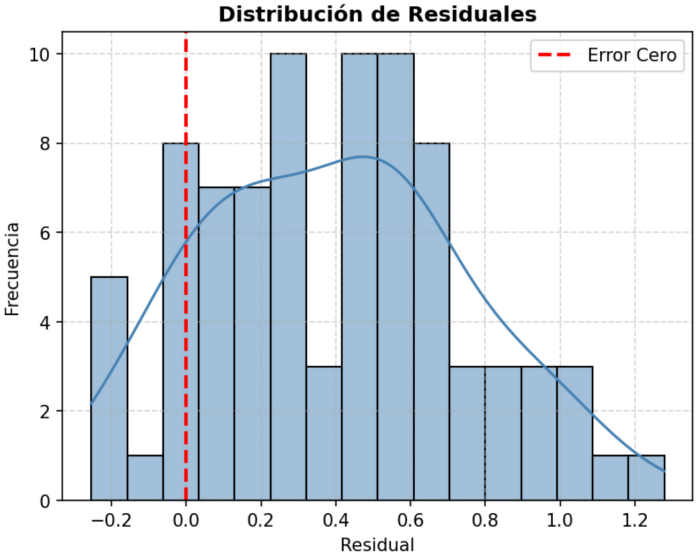

Figura 9

Interpretación de resultados

Los resultados obtenidos en el modelo son bastante acertados, el error relativo que se obtiene es bajo. En la Figura 8 podemos observar que todos los valores están acotados por $0,08$; este error no es muy alto teniendo en cuenta que hay muchas características de las viviendas que no se están considerando en el análisis de precio de las mismas, es decir, se están considerando características de ubicación geográfica de la vivienda, pero no se juzgan las características internas de la misma. Adicionalmente, podemos evidenciar en la misma figura

que la gran mayoría de errores relativos oscilan entre 0 y $0,03$; lo que claramente vendría siendo un error muy bajo y bastante por debajo de la cota. El error en nuestra aplicación específica puede costar como máximo 13000 nuevos dólares taiwaneses, que haciendo la conversión a pesos colombianos y dólares para tener una mejor percepción del valor del error obtenemos que son $1'715{,}979$ COP y 394,90 USD respectivamente. Notemos también que entre más valor por unidad de área obtenga la propiedad que estamos introduciendo a la red neuronal conseguiremos un resultado más acertado, ya que en la figura de error relativo se evidencia que los puntos se acercan en general cada vez más al error cero a medida que su valor aumenta.

Calidad del modelo y limitaciones del método

Con respecto a la calidad del modelo primero podemos notar que para realizar el entrenamiento se hicieron 75 épocas , donde como se evidencia en la Figura 7, la curva de pérdida se encuentra en uno de sus puntos más bajos, ya que entre 70 y 75 épocas es donde la pérdida se encuentra en su punto más bajo, lo que nos dice que la diferencia entre los valores reales y los valores aproximados no es significativamente grande. Un proceso que ayudó también a la buena calidad de la aplicación fue la buena elección de híper parámetros como número de neuronas de las capas ocultas, tasa de aprendizaje, tamaño del lote de datos y el número de épocas. Estos fueron escogidos en el bloque de código número 13, donde se probaron diferentes combinaciones de los parámetros y se consiguió la combinación puesta en la parte inferior del bloque.

Se puede observar en la figura 9 que en general se tiende a subestimar el valor de una unidad de área, lo cual puede incurrir en problemas a la hora de ser utilizado el modelo.

Adicionalmente, es importante considerar las limitaciones del modelo, por más que se logren encontrar híper parámetros que ayuden a que el modelo tenga una mayor precisión, este nunca logrará ir mucho más lejos de los resultados actuales porque hay muchas más variables determinantes del valor de una vivienda que las que tenemos disponibles en el conjunto de datos, tales como características internas de ella.

Posibles mejoras o diferentes alternativas

La primera alternativa para abordar la aplicación que tenemos es aplicar mínimos cuadrados directamente sobre el conjunto de datos que tenemos, obteniendo tambien una forma de predecir valores de unidad de area con una función lineal.

Una posible mejora que se ha mencionado anteriormente sería la ampliación del conjunto de datos, esto podría mejorar la capacidad del modelo de capturar la variabilidad de los precios. También se pueden crear variables calculadas derivadas de las ya existentes o las nuevas que sean incorporadas, un ejemplo sería la distancia de la vivienda al centro de la ciudad de New Taipei a partir de la latitud y longitud de la vivienda, y de este modo enriquecer la forma de juzgar las viviendas.

Además, teniendo en cuenta que el error en la predicción es mayor en valores estimados menores, se puede idear un factor de nivelación (penalización) dinámico dependiente del valor de la unidad de área predicho, donde los menores valores tengan una mayor penalización que los mayores valores resultantes del modelo.

Una estrategia adicional puede ser realizar un proceso de elección de híper parámetros mucho más riguroso o amplio, brindando más alternativas por híper parámetro o estableciendo análisis por rango, pero esto significaría arriesgarse a un mayor consumo de recursos computacionales para obtener un entrenamiento mucho más afinado y preciso. Se aclara que las mejoras que podrían obtenerse no serían significativas si se mantiene el modelo de datos tal cual como se encuentra ya que la aproximación que se obtuvo es significativamente buena como se puede ver en la figura 6.

Referencias

[1] H. Abdi, *Least Squares*, Encyclopedia for Research Methods for the Social Sciences, pp. 792–795, 2003.

[2] R.L. Burden and J.D. Faires, *Numerical Analysis*, 9th Edition, Brookscole, Boston, 2011.

[3] Strang, G. (2016). *Introduction to Linear Algebra* (5th ed.). Wellesley-Cambridge Press. Capítulos relacionados con proyección ortogonal y mínimos cuadrados.