Aprendizaje Supervisado

Encontrando la solución de regresión

ML-6561

October 31, 2023

Contenido

Índice

Regresión Lineal

La regresión lineal es un modelo estadístico que busca modelar la relación lineal entre una variable dependiente Y y una o más variables independientes X. En forma matricial, esto se representa como:

$$Y = X\beta + \varepsilon$$

donde:

- Y es el vector de variables dependientes.
- X es la matriz de variables independientes.
- β es el vector de coeficientes.
- ε es el vector de errores.

Mínimos cuadrados ordinarios (OLS)

Notación matricial

$$\min_{\beta} \|y - X\beta\|^2 \quad (1)$$

donde:

$$\boldsymbol{X} \in \mathbb{R}^{?,?}\text{Matriz de datos}$$
$$\boldsymbol{\beta} \in \mathbb{R}^{?,?} \text{ es el vector de coeficientes,}$$
$$y \in \mathbb{R}^{?,?} \text{ es el vector de observaciones}i,$$
$$m \text{ es el número de predictores}$$
$$n \text{ es el número de observaciones.}$$

Tarea

Encontrar las dimensiones de X, β y y

Mínimos cuadrados ordinarios (OLS)

Notación vectorial

$$\min_{\beta} \sum_{i=1}^{?} (y_i - \boldsymbol{x}_i^T \beta)^2 \tag{2}$$

donde:

$\boldsymbol{x}_i \in \mathbb{R}^{?,?}$ es el vector de variables predictoras para la observación i,

$\boldsymbol{\beta} \in \mathbb{R}^{?,?}$ es el vector de coeficientes,

$y_i \in \mathbb{R}^{?,?}$ es el valor observado para la observación i,

m es el número de predictores

n es el número de observaciones.

Tarea

Encontrar las dimensiones de x_i, β y y_i

Ecuaciones normales

Para encontrar los coeficientes β que minimizan el error cuadrático, se pueden utilizar las ecuaciones normales:

$$X^T X\beta = X^T Y$$

La solución para β es:

$$\beta = (X^T X)^{-1} X^T Y$$

Tarea

1. Encontrar la expresión cerrada para 1 variable.
2. Intentar conseguir la solución general.
3. Investigar la definición de $X^T X$ para cualquier matrix X.

Índice

$$\min_{\beta} \|y - X\beta\|^2 \tag{3}$$

$\boldsymbol{X} \in \mathbb{R}^{?,?}$Matriz de datos

$\boldsymbol{\beta} \in \mathbb{R}^{?,?}$ es el vector de coeficientes,

$y \in \mathbb{R}^{?,?}$ es el vector de observaciones$i$,

m es el número de predictores

n es el número de observaciones.

Tarea

Encontrar las dimensiones de X, β y y

Denotamos como $x \in \mathbb{R}^n$, un vector de n componentes.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Si por ejemplo, $n = 3$ un vector en $\mathbb{R}^3$ pudiese ser

$$x = \begin{bmatrix} 1 \\ 2 \\ 8 \end{bmatrix}$$

Podemos resolver ya para y y para β

$$\min_{\beta} \|y - X\beta\|^2 \tag{4}$$

$\boldsymbol{X} \in \mathbb{R}^{?,?}$Matriz de datos

$\boldsymbol{\beta} \in \mathbb{R}^m$ es el vector de coeficientes,

$y \in \mathbb{R}^n$ es el vector de observaciones$i$,

m es el número de predictores

n es el número de observaciones.

Ahora vamos con X

Matriz

Con $A \in \mathbb{R}^{m \times n}$ vamos a denotar matrices con m filas y n columnas, donde las entradas de A son números reales.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- m y n no tienen que ser iguales.
- Si $m = n$ entonces decimos que A es cuadrada, de lo contrario rectangular.

Podemos resolver para X

Mínimos cuadrados ordinarios (OLS)

Tarea 1

$$\min_{\beta} \|y - X\beta\|^2 \tag{5}$$

$\boldsymbol{X} \in \mathbb{R}^{n,m}$Matriz de datos

$\boldsymbol{\beta} \in \mathbb{R}^{m}$ es el vector de coeficientes,

$y \in \mathbb{R}^{n}$ es el vector de observaciones$i$,

m es el número de predictores

n es el número de observaciones.

Mínimos cuadrados ordinarios (OLS)

Tarea 2

$$\min_{\beta} \sum_{i=1}^{?} (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 \tag{6}$$

$\boldsymbol{x}_i \in \mathbb{R}^{?,?}$ es el vector de variables predictoras para la observación i,

$\boldsymbol{\beta} \in \mathbb{R}^{?,?}$ es el vector de coeficientes,

$y_i \in \mathbb{R}^{?,?}$ es el valor observado para la observación i,

m es el número de predictores

n es el número de observaciones.

Con $A \in \mathbb{R}^{m \times n}$ vamos a denotar matrices con m filas y n columnas, donde las entradas de A son números reales.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}$$

- Si queremos denotar el vector columna k, hacemos a^k.
- Si queremos denotar el vector fila k, hacemos a_k^T.

Vamos con $x_i^T\beta$

Operaciones entre vectores

- Producto interno o producto punto.

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

- Producto externo

$$xy^T \in \mathbb{R}^{m\times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

Vamos con $x_i^T\beta$

Operaciones entre vectores

En resumen, la expresión $x_i^T\beta$ es el producto interno de la fila i de X y el vector de coeficientes β, este producto nos retorna un número.

Recordemos, para dos vectores $x \in \mathbb{R}^s$ y $y \in \mathbb{R}^s$

$$x^T y \in \mathbb{R}$$

para nuestro caso es (y con esto resolvemos toda la tarea)

$$x_i^T\beta$$

$$\min_{\beta} \sum_{i=1}^{?} (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 \tag{7}$$

$\boldsymbol{x}_i \in \mathbb{R}^m$ es el vector de variables predictoras para la observación i,

$\boldsymbol{\beta} \in \mathbb{R}^m$ es el vector de coeficientes,

$y_i \in \mathbb{R}$ es el valor observado para la observación i,

m es el número de predictores

n es el número de observaciones.

$$\min_{\beta} \|y - X\beta\|^2 \tag{8}$$

$\boldsymbol{X} \in \mathbb{R}^{n,m}$Matriz de datos

$\boldsymbol{\beta} \in \mathbb{R}^{m}$ es el vector de coeficientes,

$y \in \mathbb{R}^{n}$ es el vector de observaciones$i$,

m es el número de predictores

n es el número de observaciones.

Tarea 3

Ecuaciones normales

Para encontrar los coeficientes β que minimizan el error cuadrático, se pueden utilizar las ecuaciones normales:

$$X^T X\beta = X^T Y$$

La solución para β es:

$$\beta = (X^T X)^{-1} X^T Y$$

Tarea

1. Encontrar la expresión cerrada para 1 variable.
2. Intentar conseguir la solución general.
3. Investigar la definición de $X^T X$ para cualquier matrix X.

- Si escribimos A como filas, podemos expresar Ax como,

$$y = Ax = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

- Si escribimos A por columnas, tenemos:

$$y = Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \ldots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \left[a^1\right] x_1 + \left[a^2\right] x_2 + \ldots + [a^n] \, x_n.$$

y es una combinación lineal de las columnas de A.

Curiosidad (Lo hablamos en otra clase)

Este resultado nos dice entonces que el problema fundamental de álgebra lineal $Ax = b$ se conecta con combinaciones lineales!

Es posible premultiplicar una matrix por la izquierda siempre que sea un vector fila.

- Si escribimos A por columnas, entonces podemos expresar $x^\top A$ como,

$$y^T = x^T A = x^T \left[\begin{array}{cccc} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{array} \right] = \left[\begin{array}{cccc} x^T a^1 & x^T a^2 & \cdots & x^T a^n \end{array} \right]$$

Es posible premultiplicar una matrix por la izquierda siempre que sea un vector fila.

- expresando A en términos de filas tenemos:

$$y^T = x^T A = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}$$
$$= x_1 \begin{bmatrix} - & a_1^T & - \end{bmatrix} + \ldots + x_m \begin{bmatrix} - & a_m^T & - \end{bmatrix}$$

y^T es una combinación lineal de las filas de A.

Curiosidad (Lo hablamos en otra clase)

Esto conecta el resultado fundamental de que el rango de A y de A^T es el mismo!

Tarea 3

La traspuesta

La matriz traspuesta es el resultado de "voltear" las filas y las colimnas.

Dada una matriz $A \in \mathbb{R}^{m \times n}$, su traspuesta, escrita $A^T \in \mathbb{R}^{n \times m}$, es la matriz de dimensión $n \times m$ cuyas entradas son:

$$\left(A^T\right)_{ij} = A_{ji}$$

Propiedades de la traspuesta:

- $\left(A^T\right)^T = A$
- $(AB)^T = B^T A^T$
- $(A+B)^T = A^T + B^T$

La norma de un vector $\|x\|$ es informalmente conocida como una "medida" del vector.
De manera formal, una norma es **cualquier función** $f : \mathbb{R}^n \to \mathbb{R}$ que cumpla 4 propiedades fundamentales:

1. Para todo $x \in \mathbb{R}^n, f(x) \geq 0$ (no-negatividad).
2. $f(x) = 0$ si y solo si $x = 0$ (definitividad).
3. Para todo $x \in \mathbb{R}^n, t \in \mathbb{R}, f(tx) = |t| f(x)$ (homogénea).
4. Para todo $x, y \in \mathbb{R}^n, f(x + y) \leq f(x) + f(y)$ (Desigualdad triangular).

Tarea 3

Ejemplos de Normas

La comunmente llamada norma euclideana o ℓ_2,

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \implies \|x\|_2^2 = x^T x$$

La norma 1 ℓ_1,

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

La norma infinito ℓ_∞ ,

$$\|x\|_\infty = \max_i |x_i| .$$

Asumamos las siguientes propiedades ciertas

- $\nabla_x b^T x = b$
- $\nabla_x^2 b^T x = 0$
- $\nabla_x x^T Ax = 2Ax$ (si A es simétrica)
- $\nabla_x^2 x^\top Ax = 2A$ (si A es simétrica)

Tarea 3.5

Luedo de este repaso, consigan las ecuaciones normales.

Preliminar: definición formal

- Dada una matriz con rango completo $X \in \mathbb{R}^{n\times m}$, y un vector $y \in \mathbb{R}^n$ tal que $y \notin \mathcal{R}(A)$, queremos conseguir un vector β tal que $X\beta$ esté tan cerca posible de y, **medido** a partir de la norma euclideana $\|y - X\beta\|_2^2$.
- Usando la definición $\|x\|_2^2 = x^T x$, tenemos

$$\|y - X\beta\|_2^2 = (y - X\beta)^T(y - X\beta) = \beta^T X^T X\beta - 2y^T X\beta + y^T y$$

Tarea 3
solución

- Tomando el gradiente respecto a β:

$$\begin{aligned}\nabla_\beta\left(\beta^T X^T X\beta - 2y^T X\beta + y^T y\right) &= \nabla_\beta \beta^T X^T X\beta - \nabla_\beta 2y^T X\beta + \nabla_\beta y^T y \\ &= 2X^T X\beta - 2X^T y\end{aligned}$$

- Igualando la expresión a 0 y despejando β tenemos las ecuaciones normales

$$\beta^* = \left(X^T X\right)^{-1} X^T y$$