Resumen del capítulo: Regresión lineal desde adentro

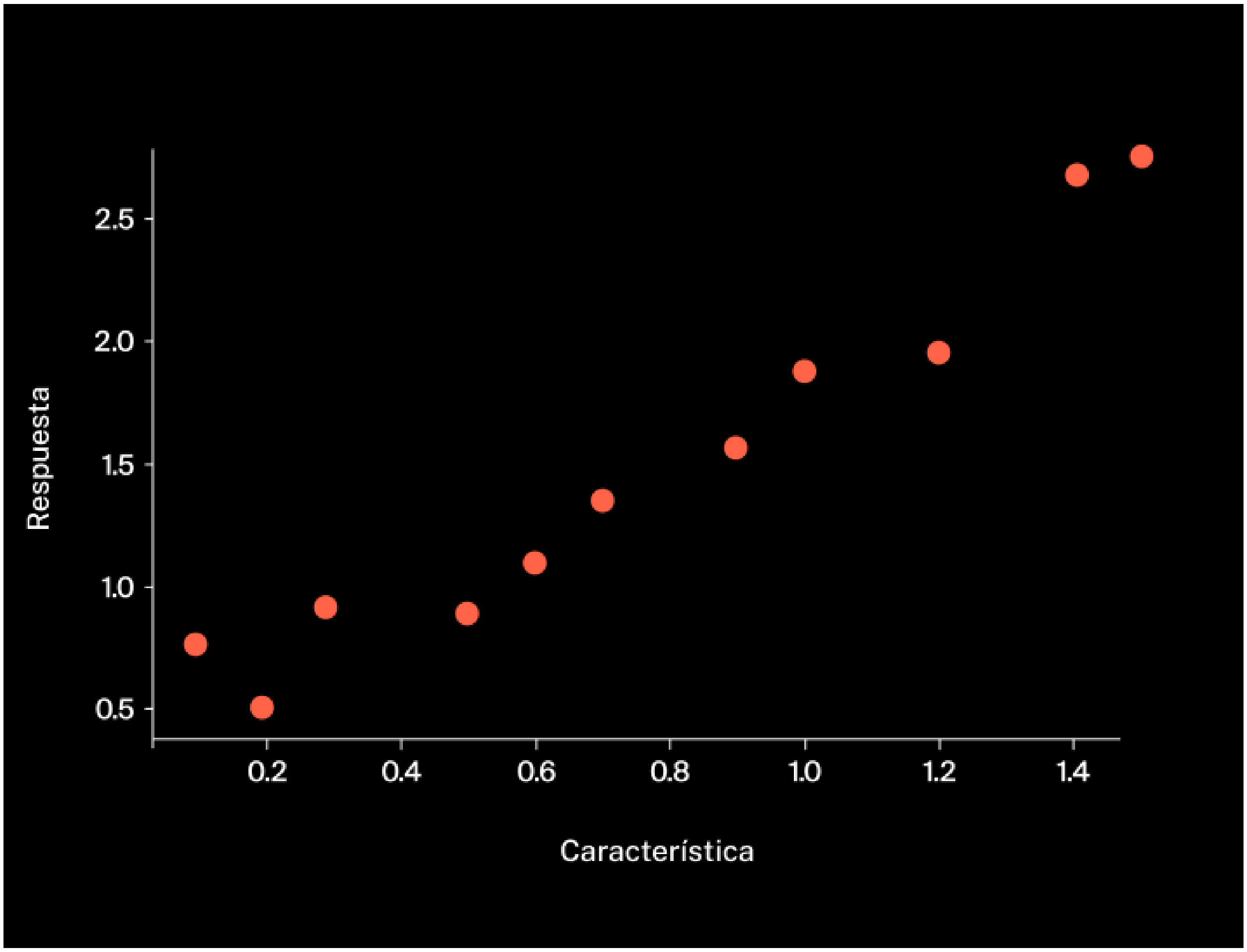
Modelo de regresión lineal

En la regresión lineal, las características son un vector de números en un espacio n-dimensional (digamos *x*). La predicción del modelo (a) se calcula de la siguiente manera: el vector de características es un escalar multiplicado por el vector de **peso** (w), luego el valor de **sesgo** de la predicción se suma a este producto:

$$a = (X \cdot w) + W_0$$

El vector w y un escalar w_0 son parámetros del modelo. Hay n parámetros en el vector w y uno en w_0. En otras palabras, el número de parámetros es mayor que la longitud del vector de características por uno.

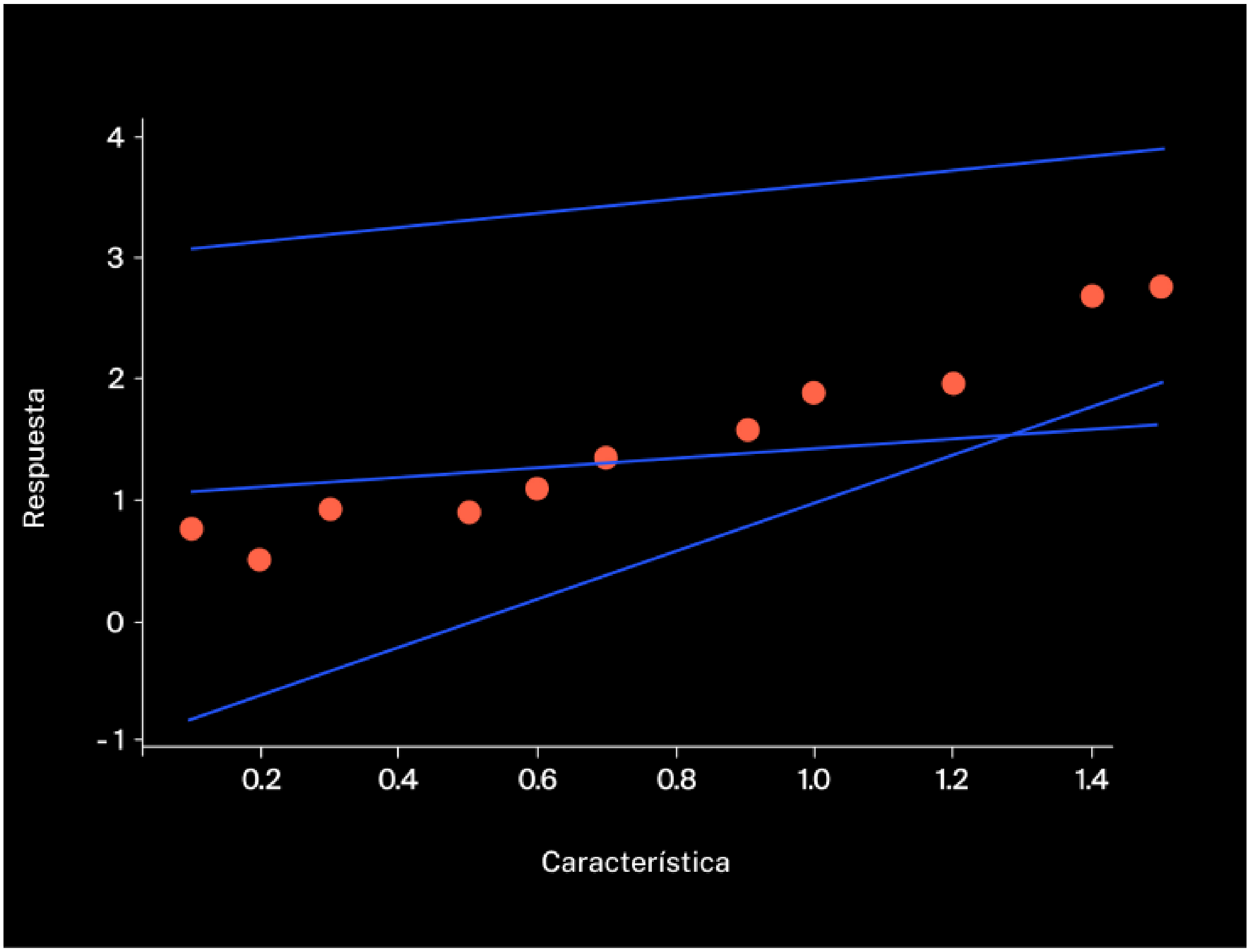
Si la longitud del vector de características es igual a uno, entonces solo hay una característica en la muestra. Dibujemos esta característica con las respuestas en el gráfico:



Los gráficos de predicción para la regresión lineal se establecen mediante la ecuación:

$$y = wx + w_0$$

Si cambias los parámetros w y w_0, obtendrás cualquier línea recta (de ahí que el modelo tome su nombre):



Objetivo de entrenamiento

Necesitamos analizar el algoritmo de aprendizaje. Nuestra métrica de calidad será ECM: el modelo debe alcanzar su valor más bajo en los datos de prueba. El objetivo de la tarea de entrenamiento se formula de la siguiente manera: encontrar los parámetros del modelo para los cuales el valor de la **función de pérdida** sea mínimo en el conjunto de entrenamiento. Como métrica de calidad, toma respuestas y predicciones correctas como entrada. Devuelve valores que representan "pérdidas" (deben minimizarse). En nuestra tarea, esta función se iguala al *ECM*. Pero, por lo general, la función de pérdida se usa para el entrenamiento mientras que la métrica de calidad se usa para las pruebas.

Vamos a escribir el objetivo de la tarea de entrenamiento en formato vectorial. El conjunto de entrenamiento se representa como matriz X, en ella, las filas corresponden a objetos y las columnas corresponden a características. Denotemos los parámetros de regresión lineal como w y w_0. Para obtener el vector de predicción a, multiplica la matriz X por el vector w y agrega el valor de sesgo de predicción w_0.

La fórmula es:

$$a = Xw + w_0$$

Para acortarla, vamos a cambiar la notación. En la matriz X, agrega una columna que consista solo en unos (será la columna 0); y el parámetro w_0 se suma al vector w:

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

$$(w_1, w_2, \dots, w_n) \longrightarrow (w_0, w_1, w_2, \dots, w_n)$$

Luego multiplica la matriz X por el vector w. El sesgo de predicción se multiplica por un vector de unos (columna cero). Obtenemos el vector de predicción resultante *a*:

$$a = Xw$$

Ahora podemos introducir una nueva notación y: el vector de valores de características objetivo para el conjunto de entrenamiento.

Escribe la fórmula para entrenar la regresión lineal de la función de pérdida del *ECM*.

$$w = \arg\min_{w} MSE(Xw, y)$$

La función `argmin()` encuentra el mínimo y devuelve los índices en los que este se alcanzó.

Matriz inversa

Una matriz identidad (también conocida como matriz unitaria) es una matriz cuadrada con unos en la diagonal principal y ceros en el resto. Si cualquier matriz A se multiplica por una matriz identidad (o viceversa), obtendremos la misma matriz A:

$$AE = EA = A$$

La **matriz inversa** para una matriz cuadrada A es una matriz A con un superíndice -1 cuyo producto con A es igual a la matriz identidad. La multiplicación se puede realizar en cualquier orden:

$$AA^{-1} = A^{-1}A = E$$

Las matrices para las que puedes encontrar inversas se llaman matrices **invertibles**. Sin embargo, no todas las matrices tienen una inversa. Esta matriz se llama matriz **no**

invertible.

Las matrices no invertibles son poco comunes. Si generas una matriz aleatoria con la función `numpy.random.normal()`, la probabilidad de obtener una matriz no invertible es cercana a cero.

Para encontrar la matriz inversa, llama a la función `numpy.linalg.inv()`. También te ayudará a verificar la invertibilidad de la matriz: si la matriz no es invertible, se detectará un error.

Entrenamiento de una regresión lineal

La tarea de entrenar la regresión lineal es:

$$w = \underset{w}{\operatorname{argmin}} \text{ ECM } (Xw, y)$$

El valor mínimo del *ECM* se obtiene cuando los pesos son iguales a este valor:

$$w = (X^T X)^{-1} X^T y$$

¿Cómo obtuvimos esta fórmula?

- La matriz de características transpuesta se multiplica por sí misma.
- Se calcula la matriz inversa a ese resultado.
- La matriz inversa se multiplica por la matriz de características transpuesta.
- El resultado se multiplica por el vector de los valores objetivo.