

Análisis predictivo con Regresión

El análisis predictivo con regresión es una técnica poderosa para predecir resultados futuros a partir de datos históricos. Esta presentación explorará los conceptos clave y las aplicaciones prácticas de este enfoque.



by Noelia Ferrero



Temario

Nociones básicas

Aprenderemos los conceptos fundamentales del análisis predictivo utilizando regresión, incluyendo cómo funciona y sus principales aplicaciones.

Aplicaciones prácticas

Exploraremos ejemplos concretos de cómo se puede aplicar el análisis predictivo con regresión en situaciones del mundo real.

Interpretación

Estudiaremos cómo interpretar los resultados del análisis predictivo, entendiendo el significado de los coeficientes y las métricas clave.

Usar el modelo

Aprenderemos a aplicar el modelo de regresión para hacer predicciones y tomar decisiones informadas.

Objetivos del encuentro



Conocer

Los conceptos de correlación, causalidad y dependencia y función lineal.



Entender

El modelo de regresión lineal y múltiple así como condiciones a cumplir.



Aplicar

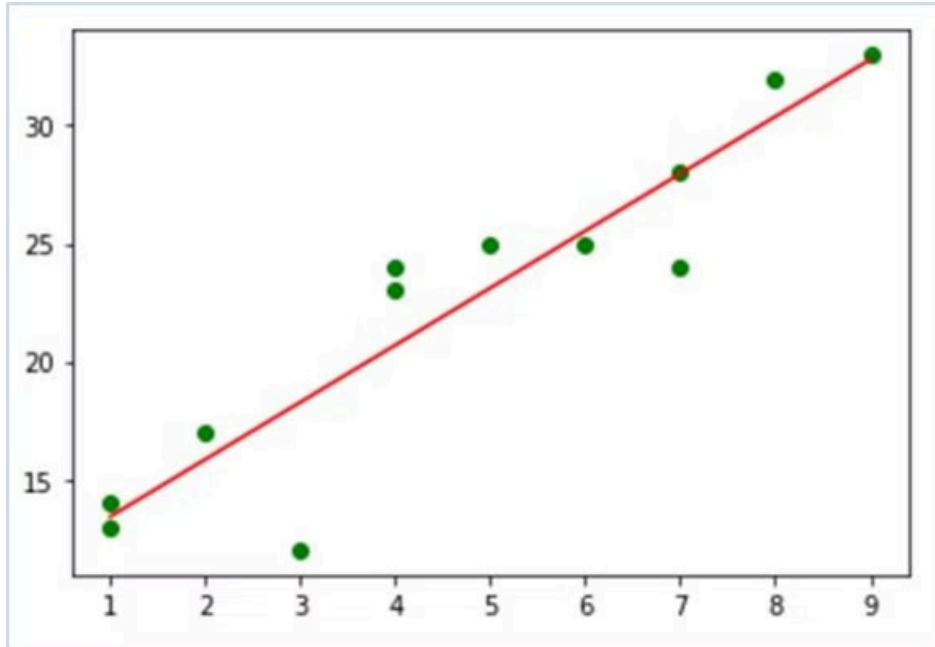
Cómo usarlo.

Datos: Temperatura vs Venta Helado

Temperatura °C	Venta Helado Kilos
12.2	125
14.4	175
15.0	325
16.7	275
17.8	425
18.3	400
19.4	450
22.2	425
22.8	450
23.9	525
25.0	550
26.7	625

y = Venta [kilos]

X = Temperatura [C]



Regresión lineal simple

En este modelo se tiene **solo una variable dependiente (y)** y **una independiente (x)**, el modelo está dado por:

$$y = \beta_0 + \beta_1 x$$

	x1	x2	x3	x4	x5	y
0	-0.524000	-1.483389	-0.406411	0.493357	-1.555619	-231.157272
1	1.029961	-0.074028	0.575738	-0.450641	0.311754	64.055904
2	-1.008525	0.874473	-0.155143	0.439377	2.369878	217.896331
3	-1.197588	0.779590	0.719375	0.847185	-0.292350	37.873532
4	0.878875	1.302094	-0.043368	-0.065888	-0.687116	68.114257
...
495	-0.138358	1.754984	0.954566	-1.375073	-0.972465	-26.421356
496	0.405534	0.389499	1.636640	-1.634547	1.922671	153.438469
497	-1.689103	0.141811	0.709097	0.606084	-0.507516	-60.723908
498	-0.241072	0.603173	-0.487081	1.037277	0.737680	133.614266
499	0.878752	0.759294	0.174286	-1.098081	-0.595734	-14.868572

500 rows × 6 columns

Regresión lineal múltiple

Es una generalización del modelo de regresión lineal simple

Para este caso tenemos una variable dependiente **y** con **M** variables independientes:

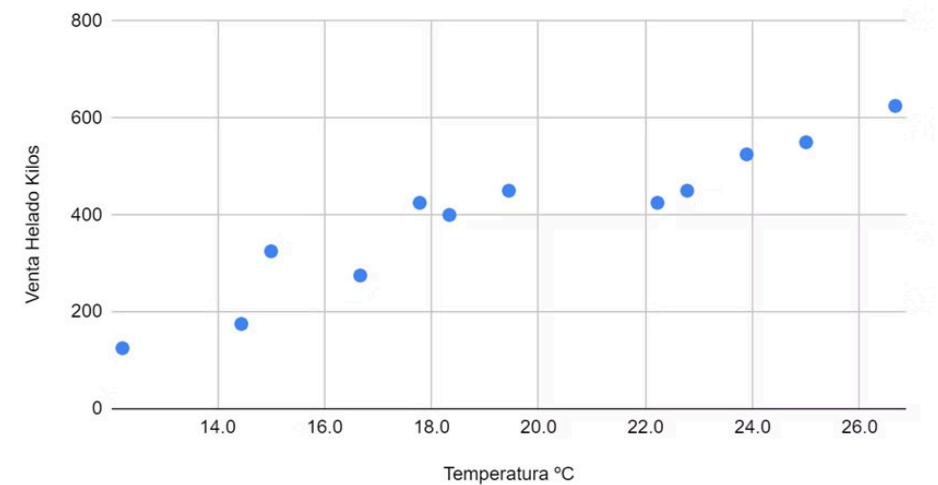
$$x_1, x_2, x_3, \dots, x_M$$

Llamaremos a las variables **x** **regresores**

Visualización

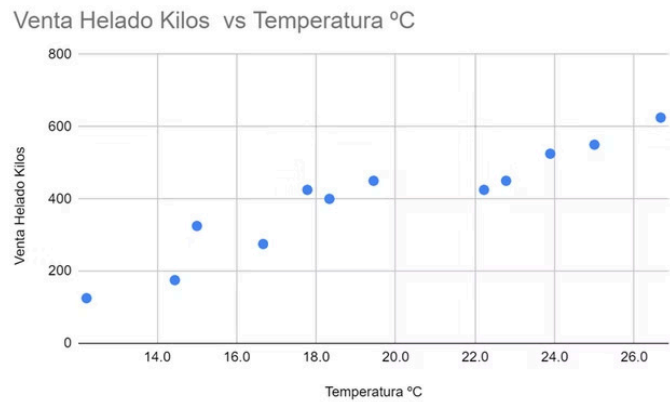
En esta sección, exploraremos la visualización de los datos de temperatura y venta de helados. La visualización nos ayudará a comprender mejor la relación entre estas dos variables y a identificar patrones y tendencias que puedan ser útiles para el análisis predictivo.

Venta Helado Kilos vs Temperatura °C



Modelado

$$y = f(X)$$



Pero, ¿Quién es f ?

Modelado: Continuación

Ahora que hemos introducido los conceptos básicos del modelado, vamos a profundizar más en el proceso. Comenzaremos explorando cómo seleccionar las variables predictoras adecuadas y cómo evaluar la calidad del modelo. También discutiremos cómo interpretar los coeficientes del modelo y cómo utilizar el modelo para hacer predicciones.

Es importante tener en cuenta que el modelado es un proceso iterativo. A menudo, tendremos que probar diferentes combinaciones de variables predictoras y ajustar el modelo hasta que encontremos uno que se ajuste bien a nuestros datos y cumpla con nuestros objetivos. La clave es mantener una mente abierta y estar dispuestos a experimentar.

Como una función matemática estándar

Donde la variable **y** es una función de **x**, o sea que **en definitiva y depende del cambio de x**. Otra forma de decir lo mismo es que **x** es una variable independiente, o sea que su cambio no depende de nuestro modelo.



Modelado: Continuación

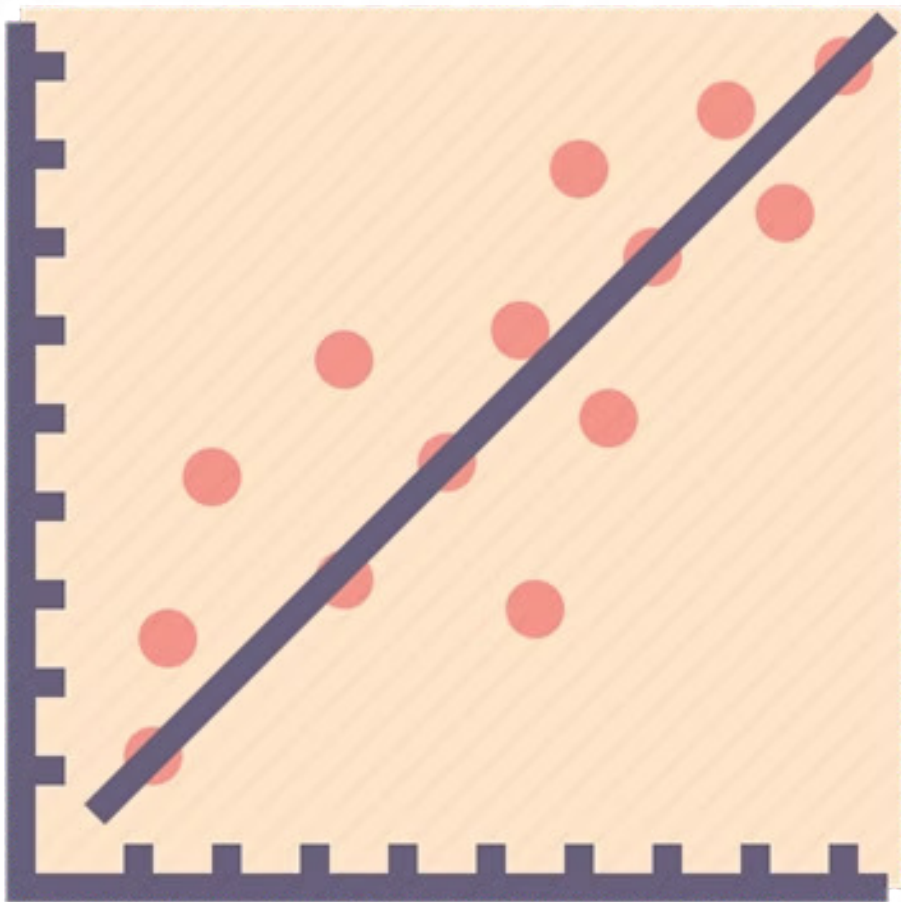
1. Continuamos explorando el proceso de modelado de la relación entre la temperatura y las ventas de helado.
2. Utilizaremos técnicas estadísticas para encontrar la función matemática que mejor se ajuste a los datos observados.
3. Analizaremos los supuestos y las limitaciones del modelo seleccionado, asegurándonos de que sea adecuado para nuestros propósitos.



Modelado: Conclusión

En conclusión, el proceso de modelado es fundamental para comprender y predecir el comportamiento de los datos. Hemos explorado cómo la regresión lineal nos permite establecer una función matemática que relaciona la temperatura con las ventas de helados. Esta función nos permite hacer predicciones sobre las ventas futuras en función de la temperatura. Es importante tener en cuenta que el modelo es una simplificación de la realidad y que pueden existir otros factores que influyen en las ventas, por lo que el modelo debe ser validado y ajustado continuamente a medida que se disponga de más datos. Luego de ese paso, discutiremos cómo evaluar la calidad del modelo y cómo utilizarlo para hacer predicciones.

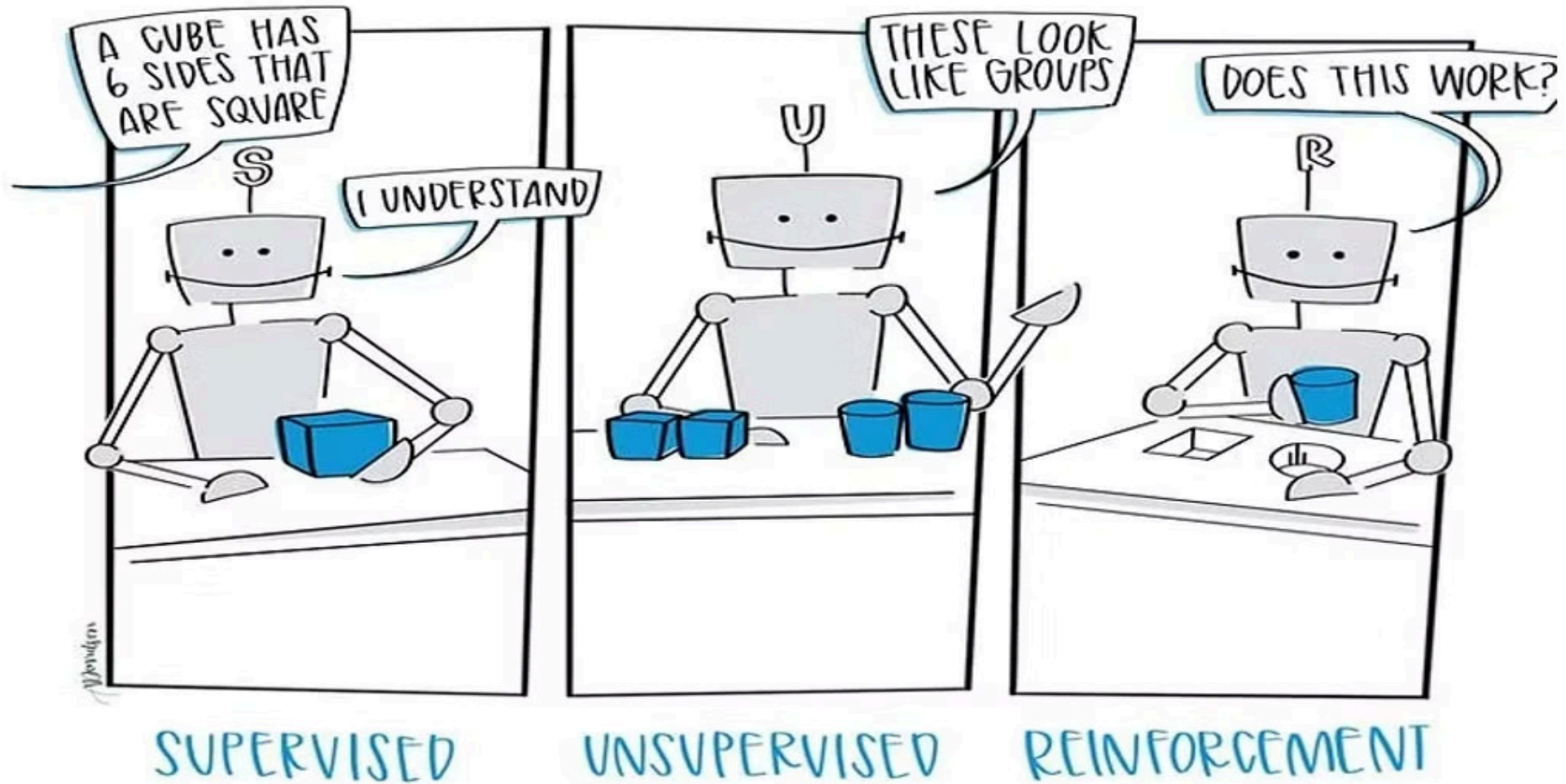




A tener en cuenta:

- ✓ Una **recta totalmente horizontal** > tiene una **pendiente igual a cero**.
- ✓ Una **recta inclinada** en el sentido de la correlación positiva > tiene una **pendiente positiva**.
- ✓ Una **recta inclinada** en el sentido de la correlación negativa > tiene una **pendiente negativa**.
- ✓ Una **recta vertical** > tiene pendiente **infinita**.

MACHINE LEARNING



Three classes of learning problems

Supervised Learning

Data: (x, y)

x is an input data, y is a label
(e.g. photo with label "cat")

Goal: Learn to map input to output
i.e. $x \rightarrow y$

An example: to classify



This is a cat

Unsupervised Learning

Data: x

x is data, there's no labels!

Goal: Learn an underlying structure
of the data.

An example: Comparison



The two things are alike

Reinforcement Learning

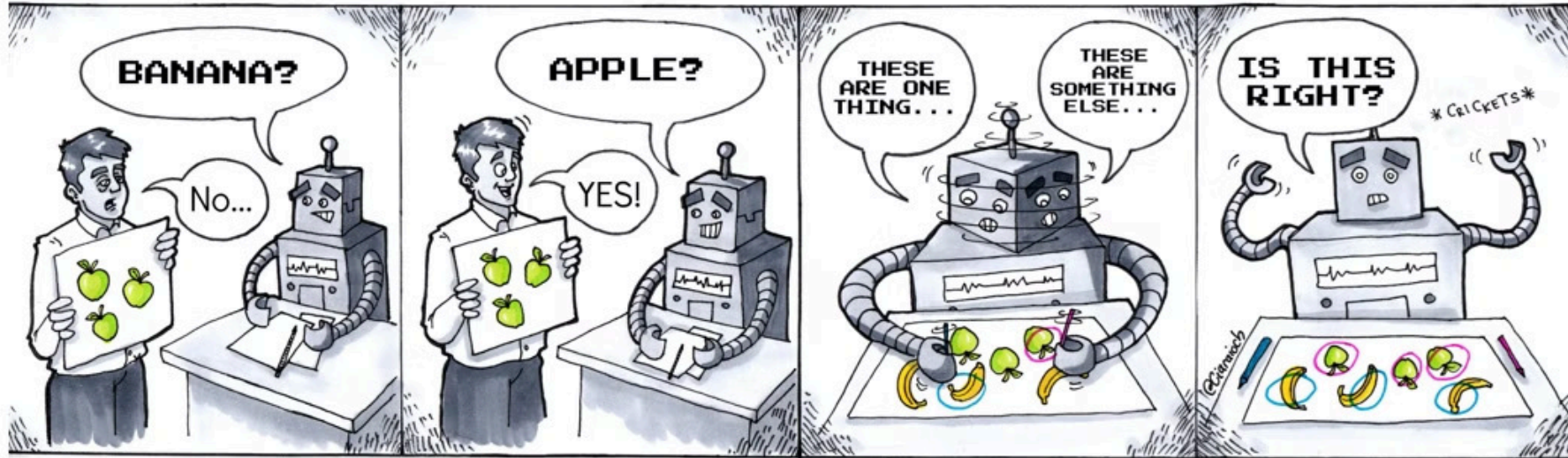
Data: No data, Only state-action
pairs (s, a) .

Goal: Maximize future reward over
many time steps.

An example: reward = joy



Interaction with the cat
gives joy



Supervised Learning

Unsupervised Learning

scikit-learn algorithm cheat-sheet

