

# Aprendizaje automático

## Modelos de regresión no lineal

¿Qué hay más allá de la  
regression lineal?

---

# Regresión polinómica

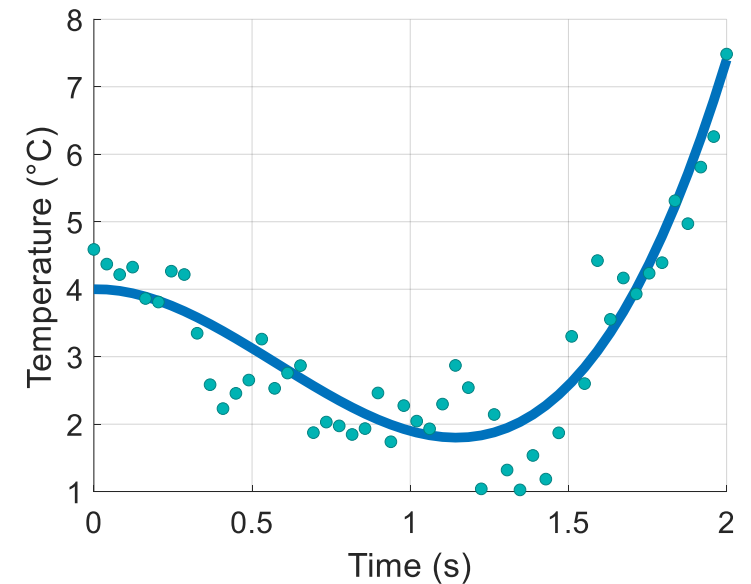
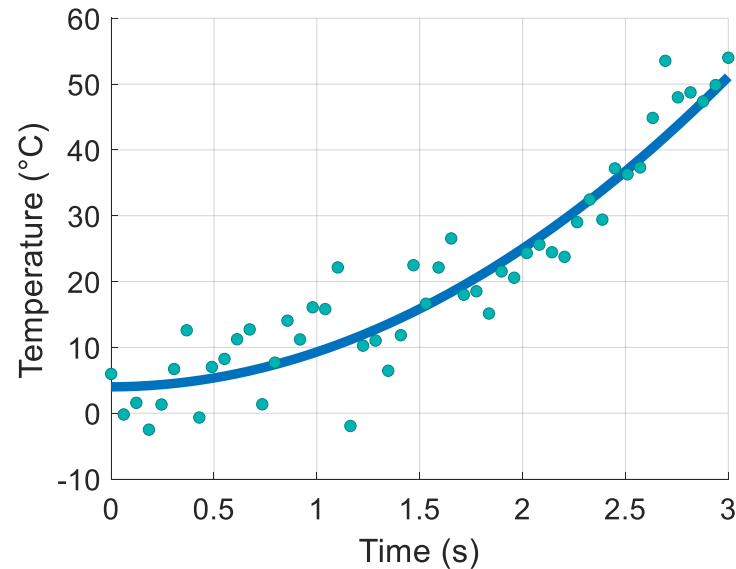
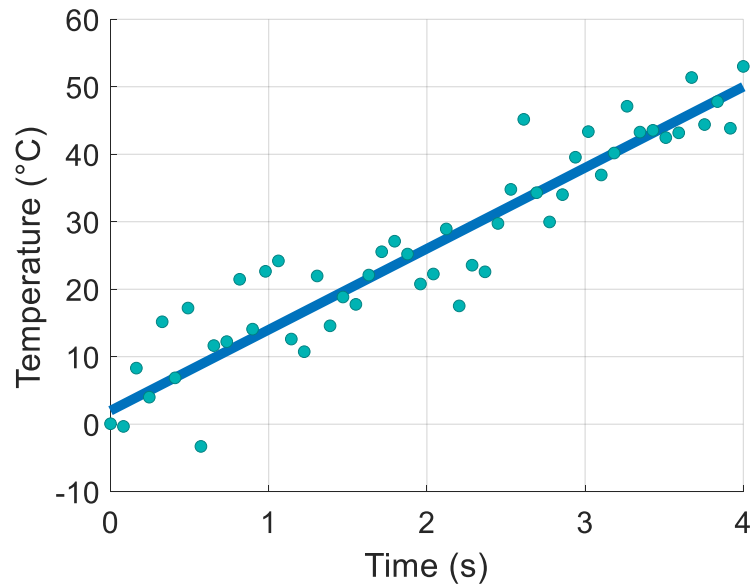
- Podemos agregar nuevas variables predictoras al modelo que sean transformaciones de las variables originales.
- Por ejemplo, si tenemos variables  $x_1$  y  $x_2$ , podemos generar el modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \beta_6 x_1^2 + \beta_6 x_2^3$$

- Con el proceso de selección de variables podemos descartar términos del modelo.

# Selección de la función de regresión

- Graficar los datos puede ayudarnos a determinar qué función puede ser apropiada para un problema dado.



---

En teoría, podemos hacer cualquier transformación de las variables predictoras para generar nuevas variables.

Esto no cambia el esquema general de la regresión lineal, ni los métodos para el ajuste y análisis de los parámetros del modelo.

---

# Regresión lineal generalizada

- Además de cambiar los predictores del modelo, podemos cambiar la relación que hay entre la variable  $y$  y  $X\beta$ .

- En el caso de regresión lineal, tenemos que:

$$y = X\beta$$

- En regresión lineal generalizada:

$$y = g(X\beta)$$

donde  $g(u)$  es una función de enlace.

# Modelos de regresión lineal generalizada

| Función de enlace   | Valores de la variable $y$     | Usos  |
|---|--------------------------------|---|
| Identidad<br>$g(\mu) = \mu$                                 | Reales de $-\infty$ a $\infty$ | Respuesta lineal de los datos.  |
| Negativa inversa<br>$g(\mu) = -\frac{1}{\mu}$               | Reales de $0$ a $\infty$       | Respuesta exponencial, parámetros de escala.                          |
| Cuadrada inversa<br>$g(\mu) = \frac{1}{\mu^2}$              | Reales de $0$ a $\infty$       |   |
| Logarítmica<br>$g(\mu) = \ln(\mu)$                          | Enteros: $0, 1, 2, 3, \dots$   | Conteo de ocurrencias en un intervalo fijo.                           |
| Logística<br>$g(\mu) = \ln\left(\frac{\mu}{1 - \mu}\right)$ | Entero: $0, 1$                 | La salida sólo tiene dos posibles valores.                            |
| Logística<br>$g(\mu) = \ln\left(\frac{\mu}{N - \mu}\right)$ | Enteros: $0, 1, 2, \dots, N$   | Conteo de número de respuestas positivas de un total de $N$ intentos. |

# K vecinos más cercanos

- Para un valor de  $X$ , se buscan las  $k$  observaciones más cercanas y se promedian los correspondientes valores de  $y$ .

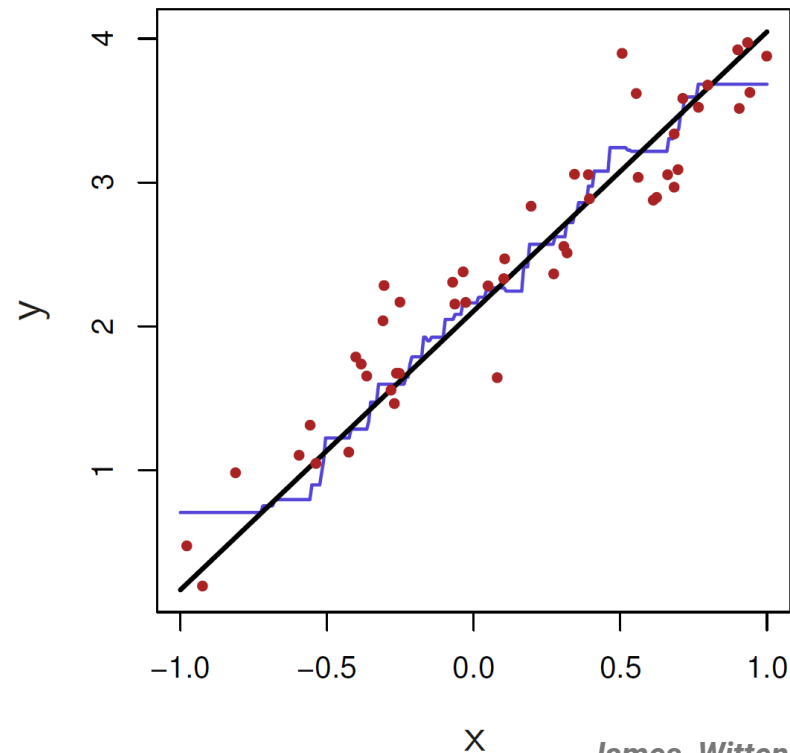
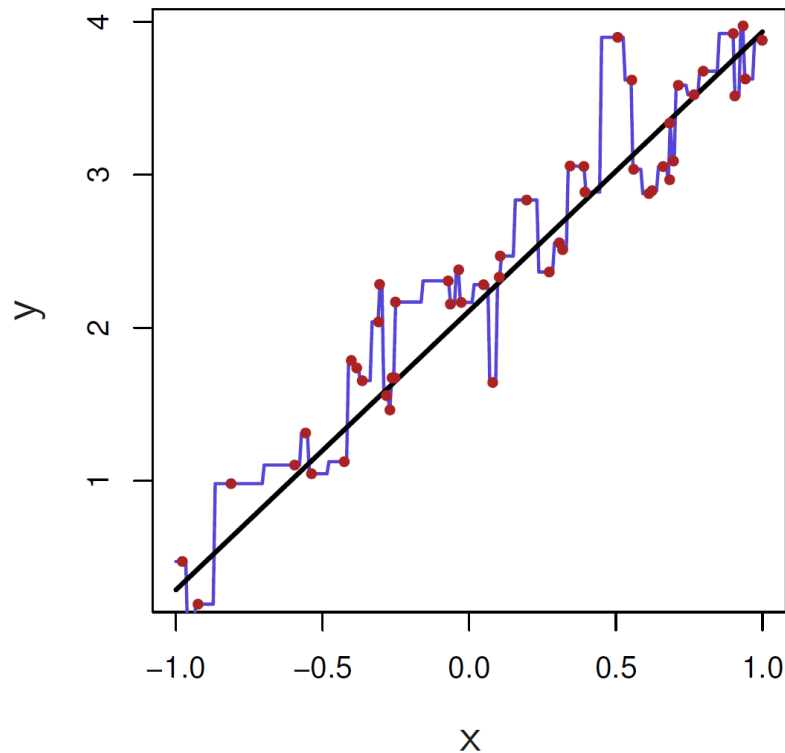
$$\hat{y} = \frac{1}{k} \sum_{i \in \mathcal{N}_X} y_i$$

donde  $\mathcal{N}_X$  denota el conjunto de los índices de los  $k$  vecinos más cercanos de  $X$ .



# K vecinos más cercanos

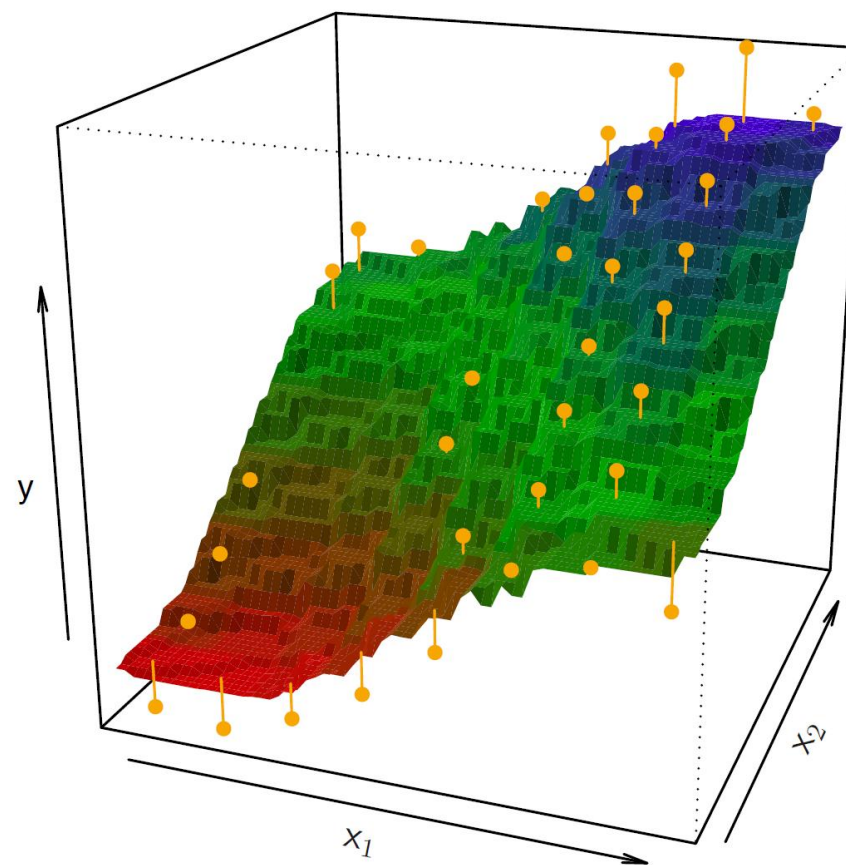
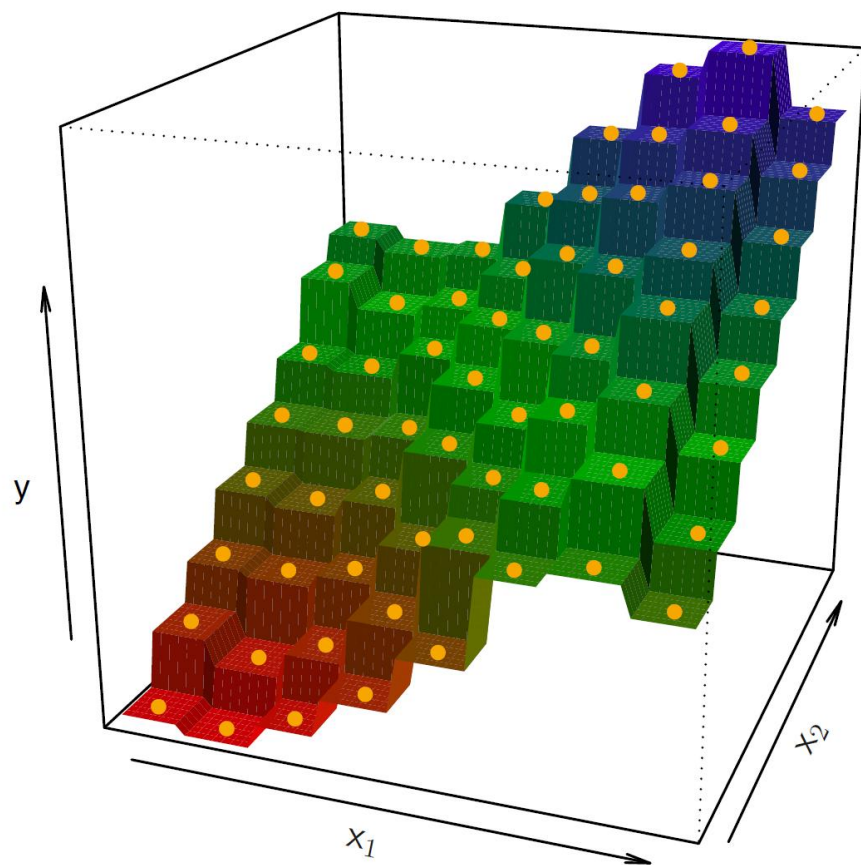
- El parámetro  $k$  controla el nivel de suavizado de la función. Para  $k$  pequeñas, la función tiende a tener cambios bruscos.



James, Witten, Hastie, & Tibshirani, 2021



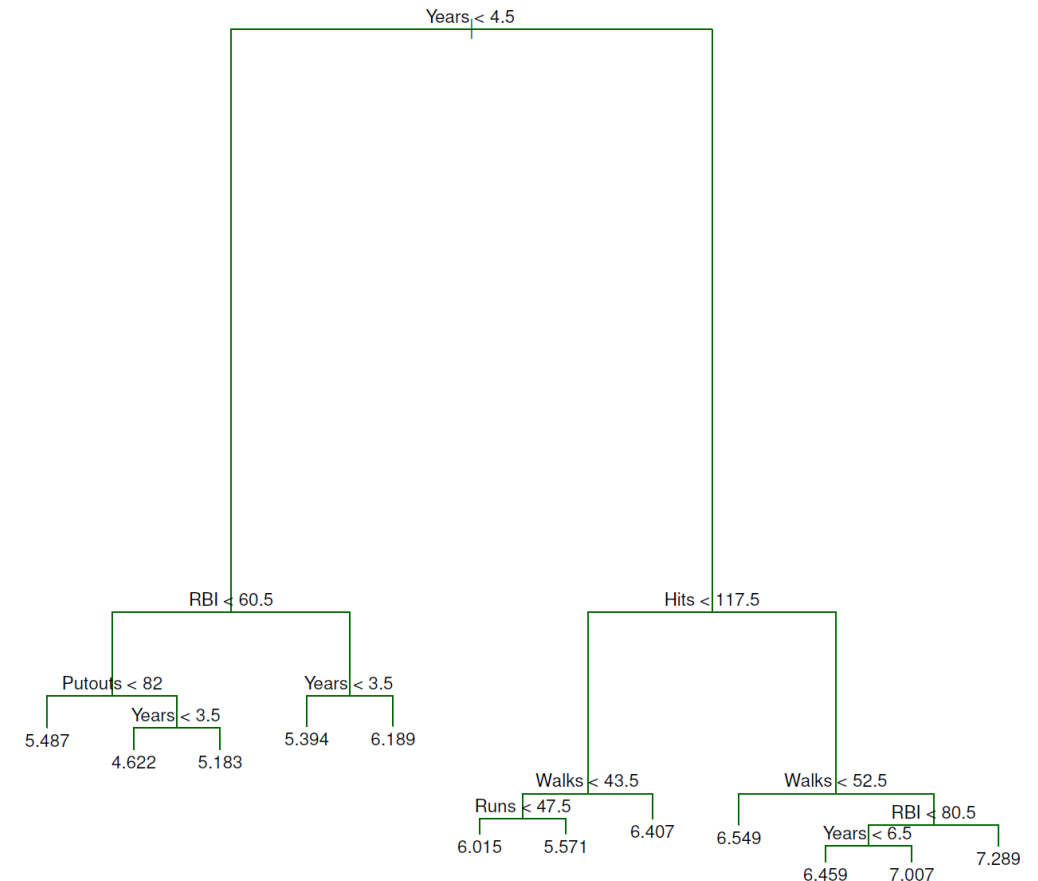
# K vecinos más cercanos



James, Witten, Hastie, & Tibshirani, 2021

# Árboles de decisión

- Los **árboles de decisión** son un conjunto de reglas que se establecen para las variables predictoras, de tal manera que el valor de la variable dependiente se obtiene al final de verificar cada una de las reglas del árbol.
- Cada nodo del árbol es una pregunta que se hace sobre una variable independiente, mientras que las hojas del árbol son los valores que se asignarían a cada caso a la variable  $y$ .



## Otros métodos de regresión

- Regresión con kernels
- Regresión local
- Splines de suavizamiento
- Regresión de vectores de soporte
- Redes neuronales



¿Qué modelo debería utilizar?

---

No existe una respuesta única a la pregunta sobre qué modelo es mejor de manera general.

Para encontrar un modelo adecuado, sólo se puede probar diferentes alternativas para el conjunto de datos dado utilizando las metodologías de validación y prueba apropiados.

---

---

En optimización y aprendizaje automático, existen los teoremas **No Free Lunch**, los cuales establecen que, si un método de optimización o modelo tiene alguna mejora para un tipo de problema, se compensa con un rendimiento menor en otra clase de problema.

---

# Bibliografía

- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2023). *An introduction to statistical learning: with applications in Python* (2da ed.). Springer.
  - Capítulo 7
  - Capítulo 8