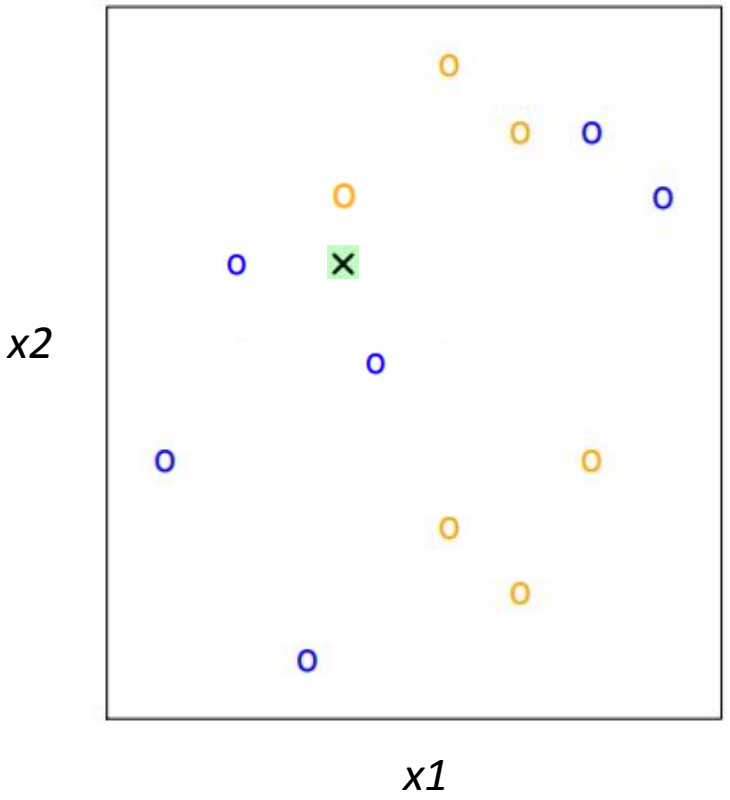
ITBA
Instituto Tecnológicc
de Buenos Aires

11/OCTUBRE

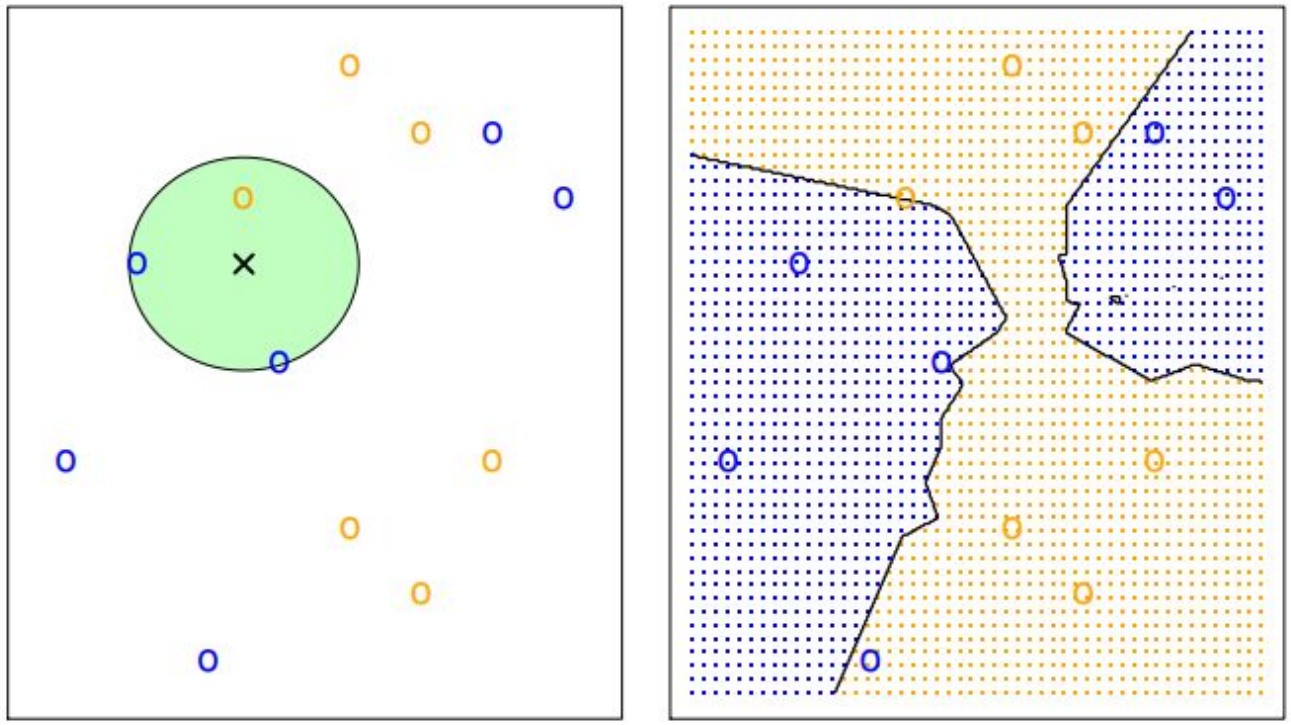
K-NEAREST NEIGHBOURS



¿Cómo predecimos la clase de **x**?

Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

K-nearest Neighbors (KNN)



K=3

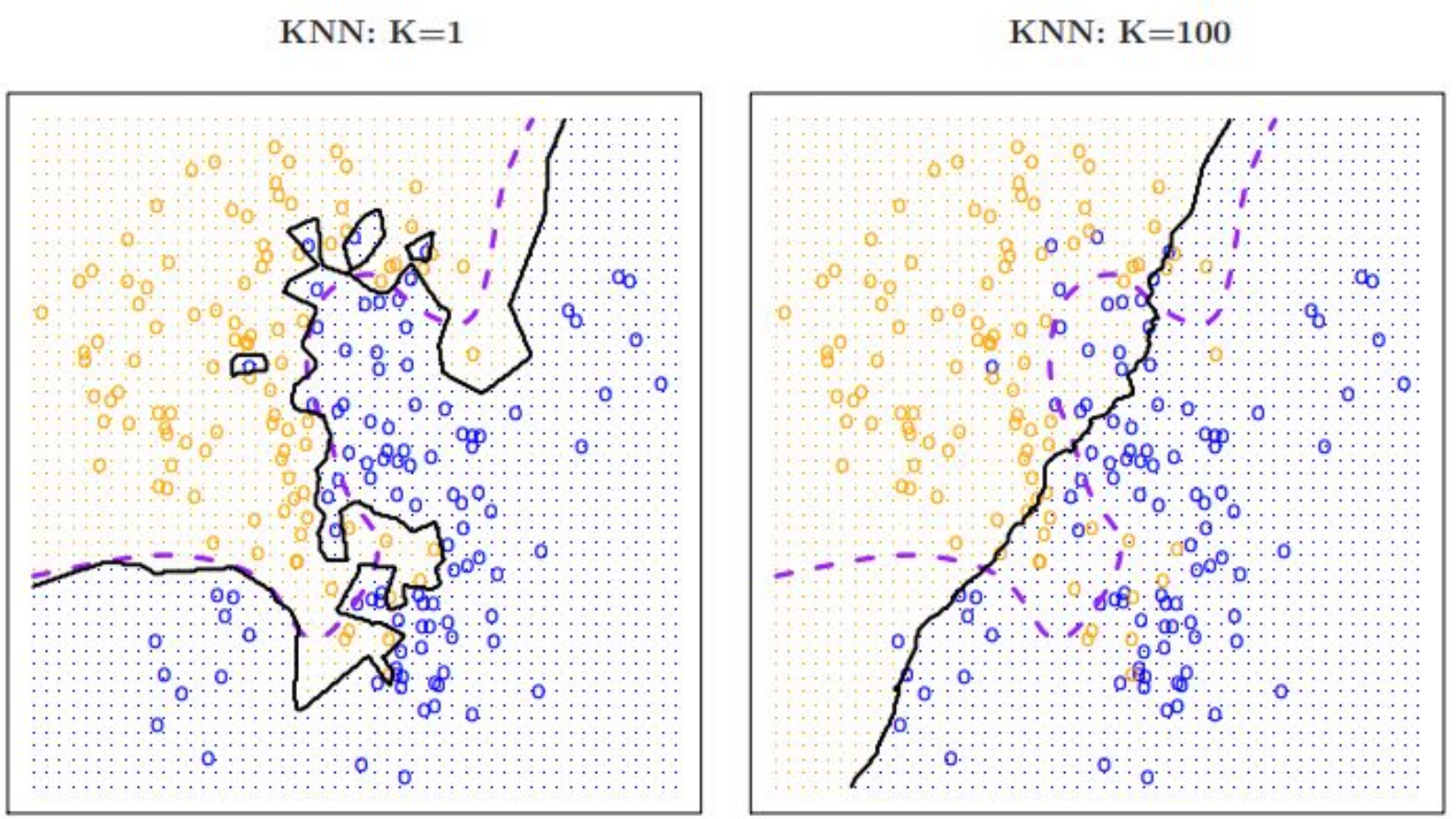
$$\Pr(Y = j|X = x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} I(y_i = j)$$

Para clasificar x_0 :

1. Buscar *K* puntos en train más cercanos a x_0 [***vecindario N_0***]
2. Calcular la distribución de clases en N_0 [***Pr(Y)***]
3. Asignar a la clase de mayor probabilidad (solemos usar *K* impar)

Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN



- - *Componente sistemático del DGP ("verdadera" frontera de clasificación)*

o o *Observaciones de entrenamiento de cada clase*

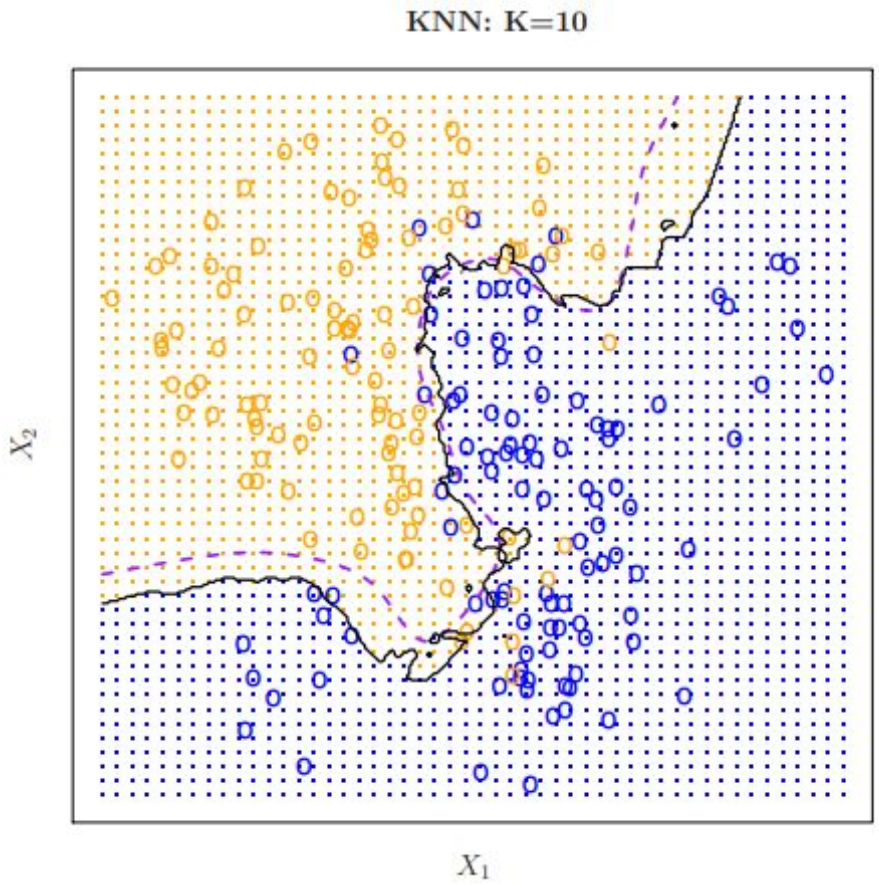
Tasa de error de clasificación

$$\frac{1}{n}\sum_{i=1}^{n} I(y_i \neq \hat{y}_i)$$

¿Cómo es el error en cada *K*?

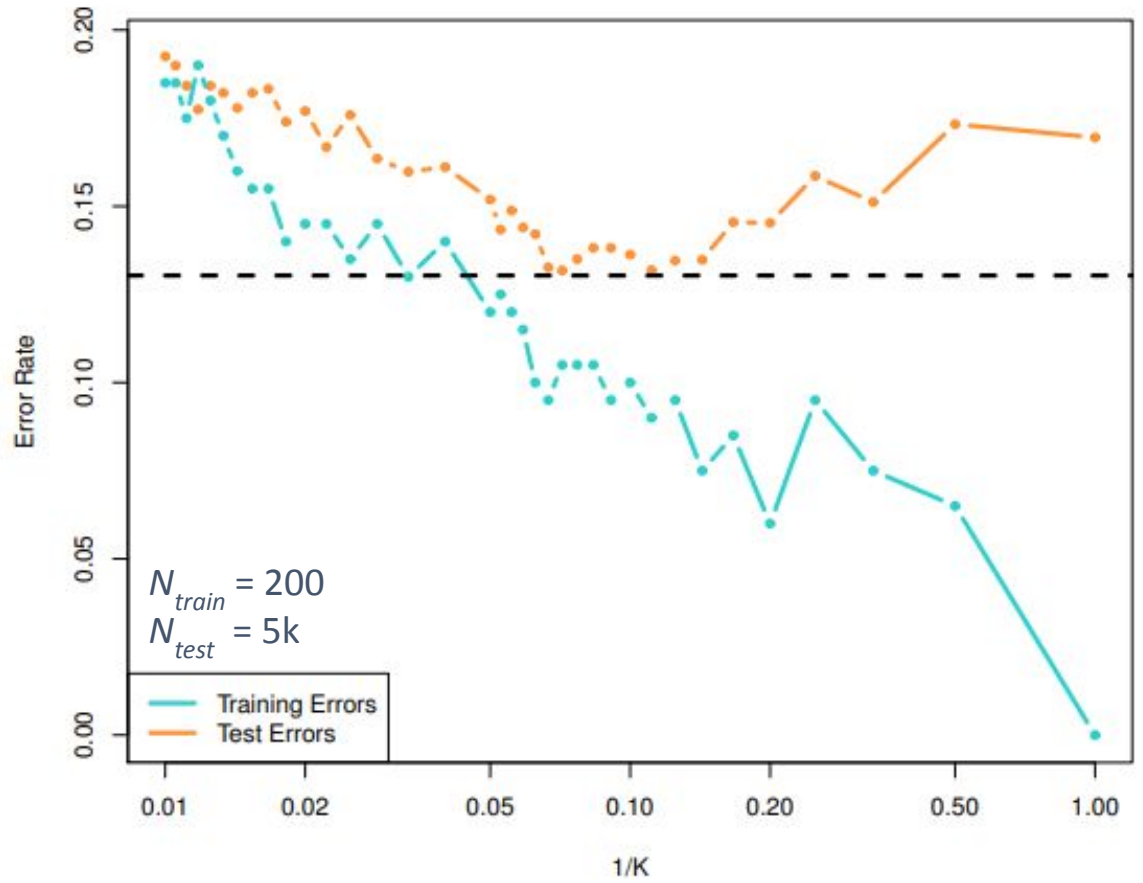
Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN



Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

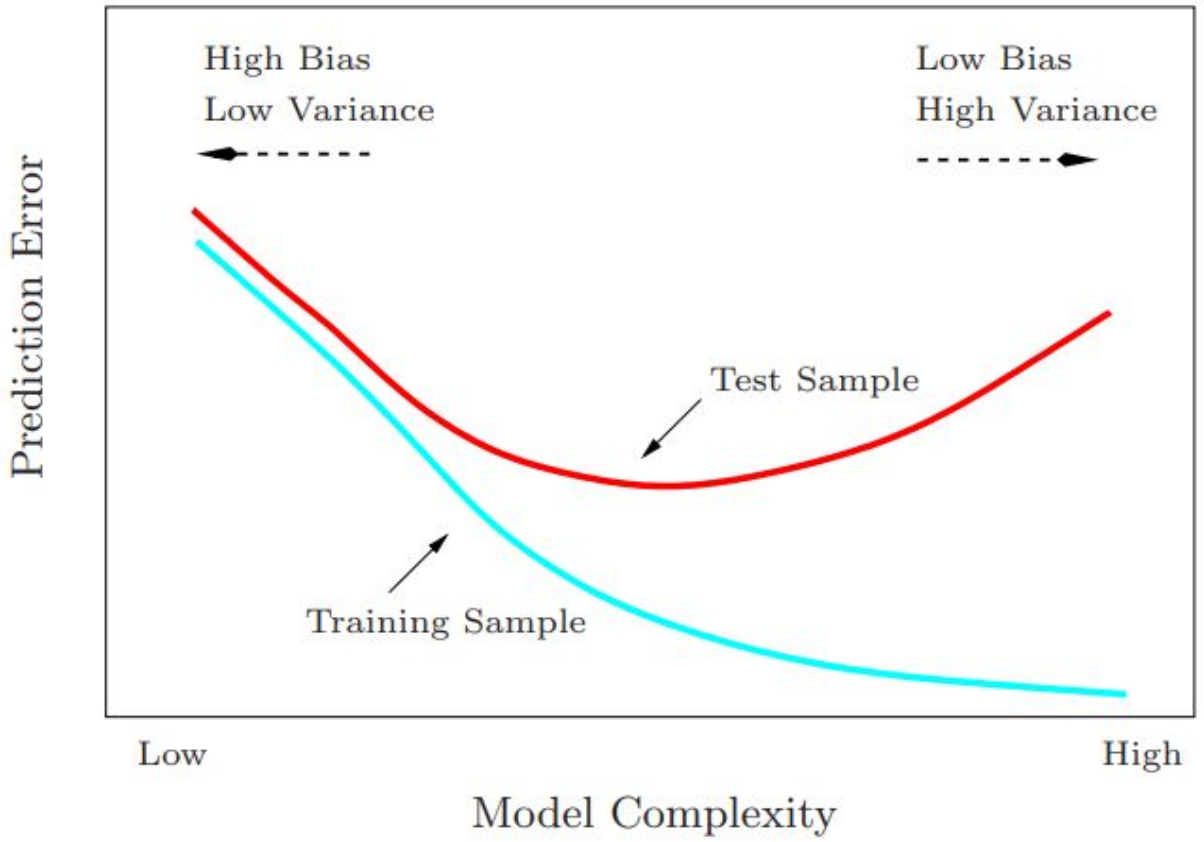
KNN



error irreducible

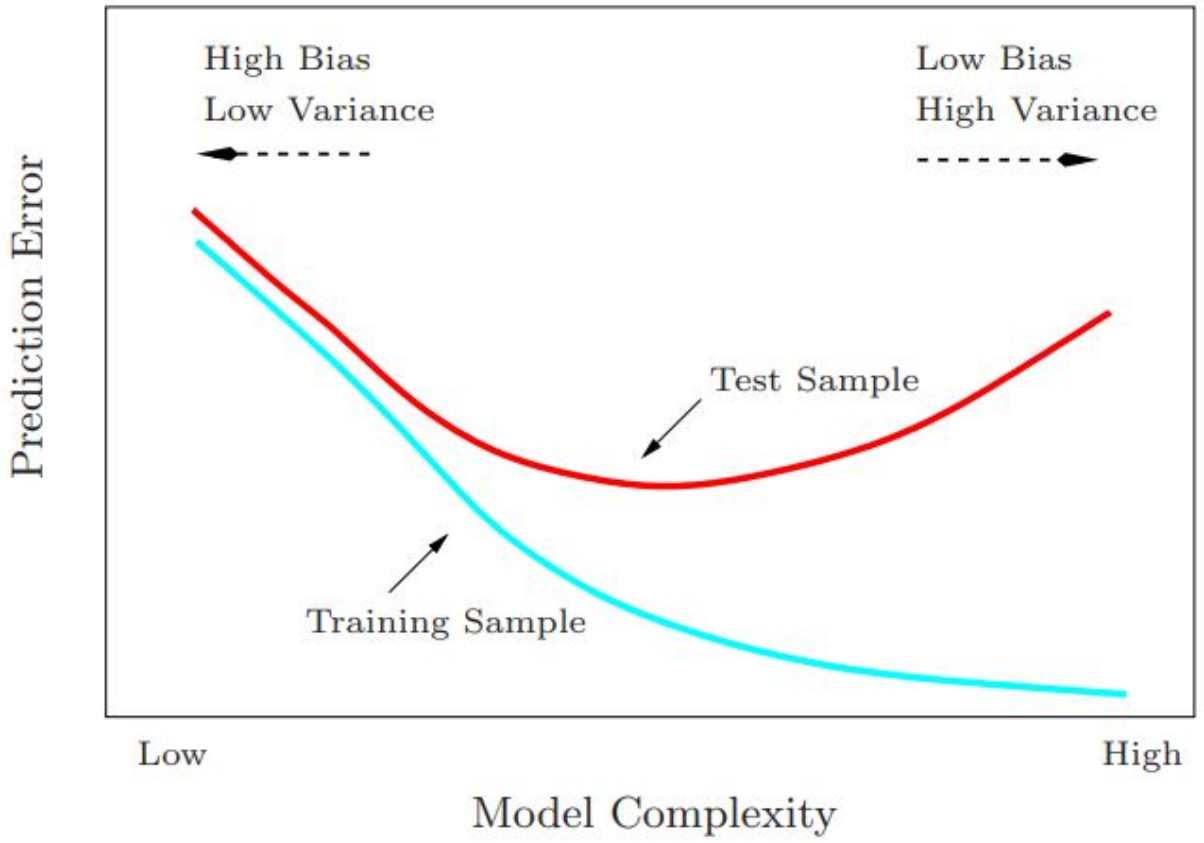
Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

Tradeoff sesgo-varianza



Fuente: The Elements of Statistical Learning (Hastie *et al*, 2009)

Tradeoff sesgo-varianza



+k -flexibilidad
⟶ -sesgo +varianza

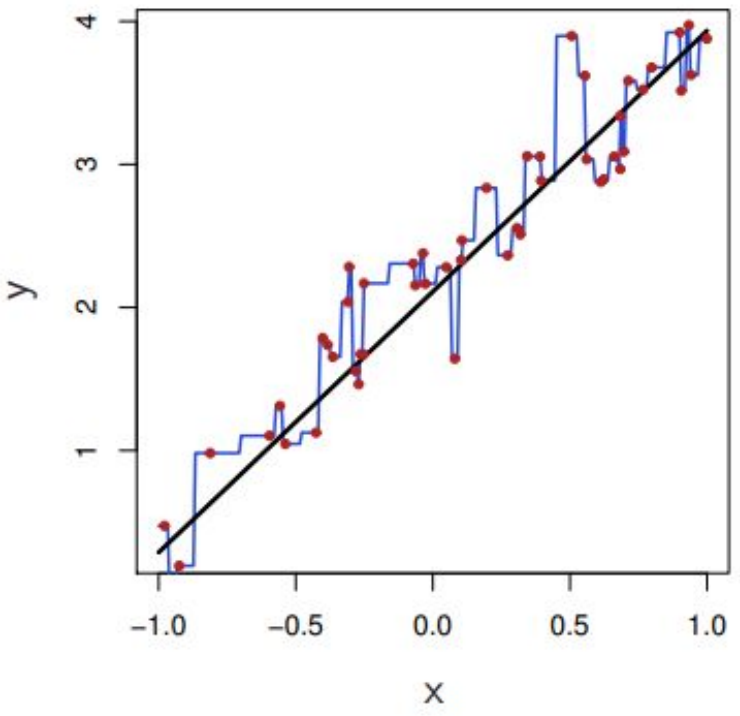
-k -flexibilidad
⟶ +sesgo -varianza

Para *K=1*, la predicción depende de una sola observación!
Cuando *K* crece, depende del promedio de muchas

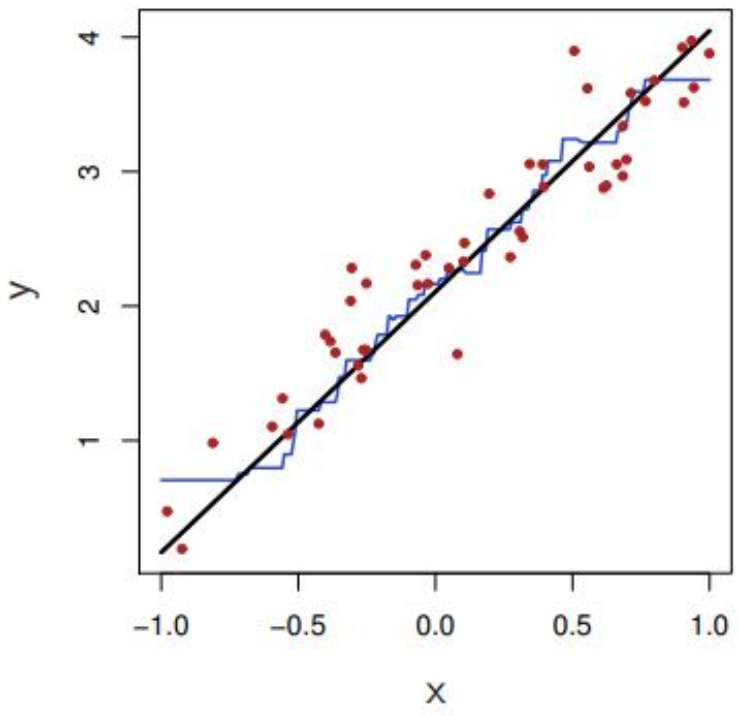
Fuente: The Elements of Statistical Learning (Hastie *et al*, 2009)

KNN regression

K=1



K=9



$$\hat{f}(x_0) = \frac{1}{K} \sum_{x_i \in \mathcal{N}_0} y_i$$

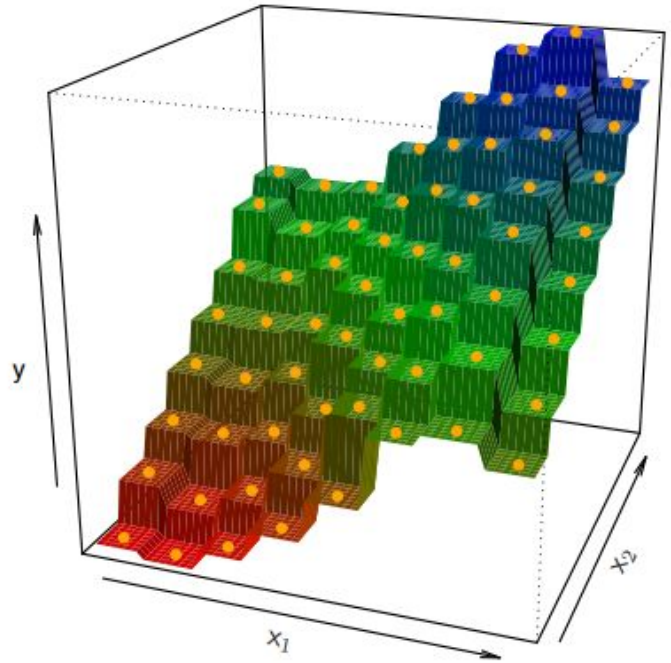
Para predecir la respuesta de x_0 :

1. Buscar *K* puntos en train más cercanos a x_0 [***vecindario*** $\boldsymbol{N_0}$]
2. Calcular el promedio de *Y* en N_0 [***f_hat(Y)***]

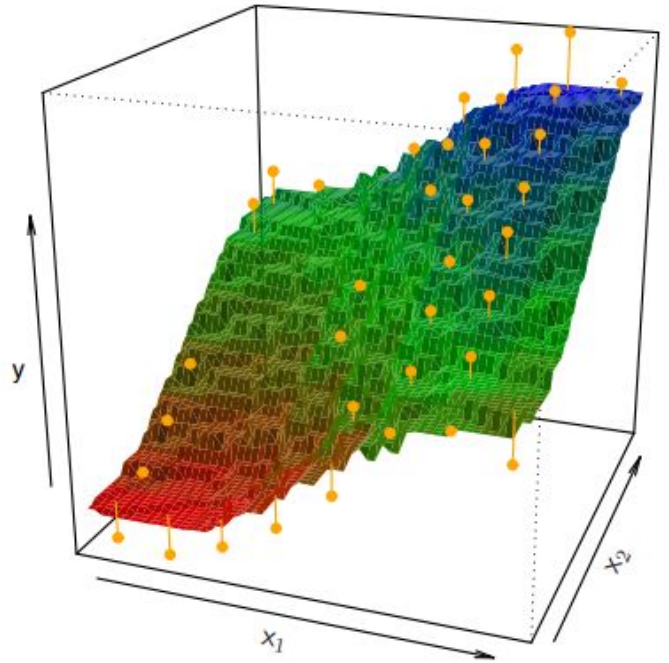
Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN regression

K=1

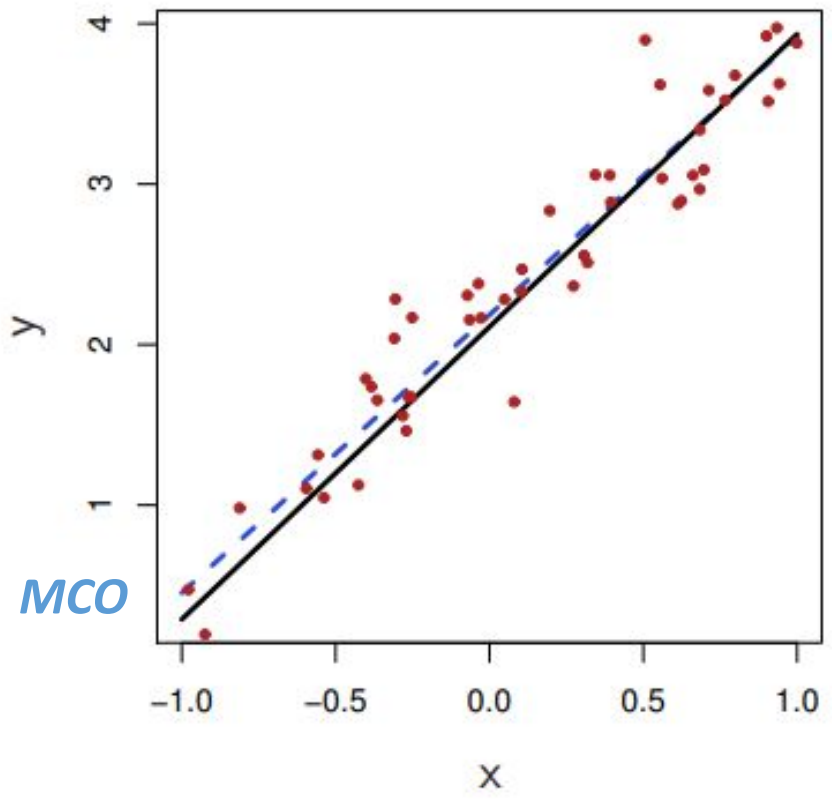


K=9

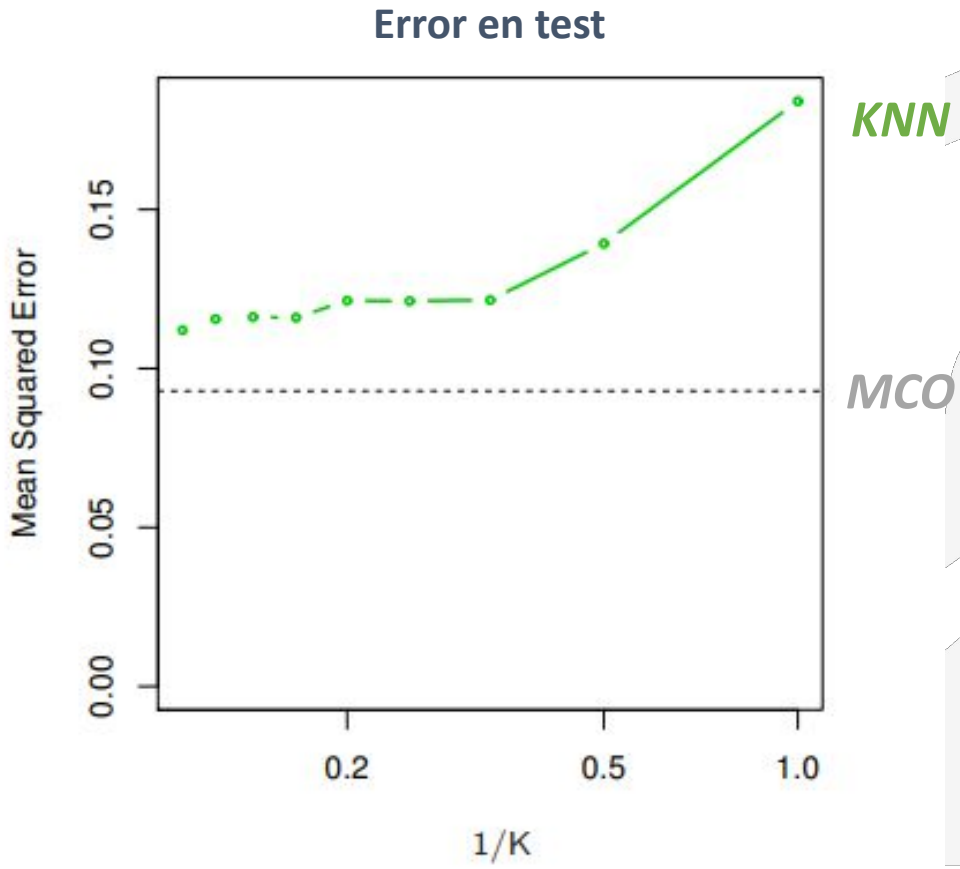


Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN regression



MCO

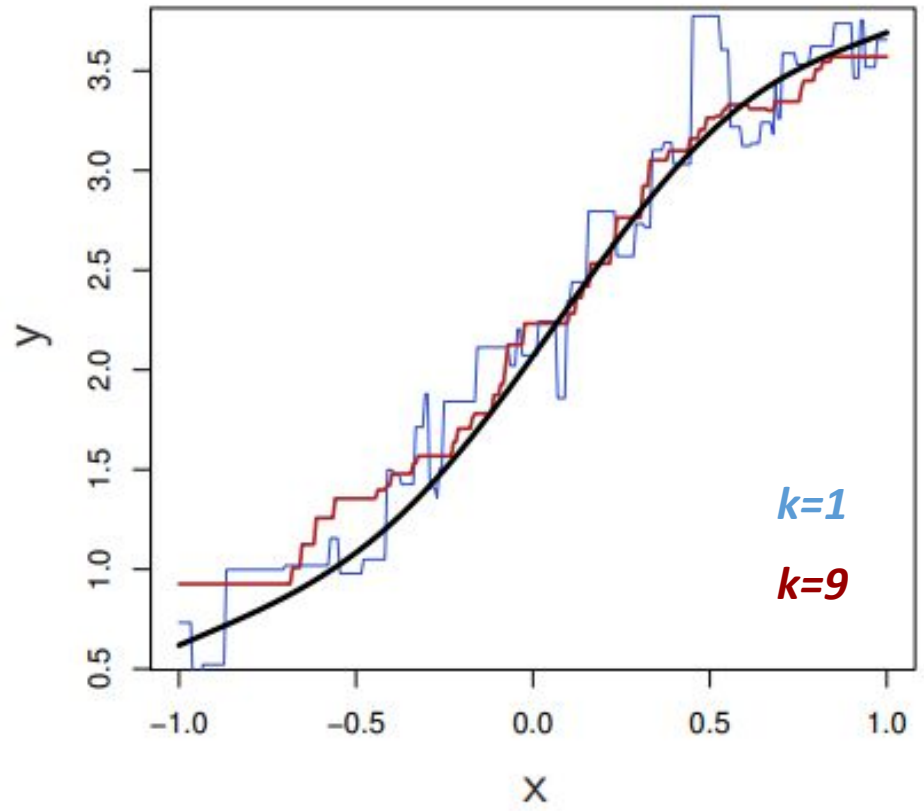


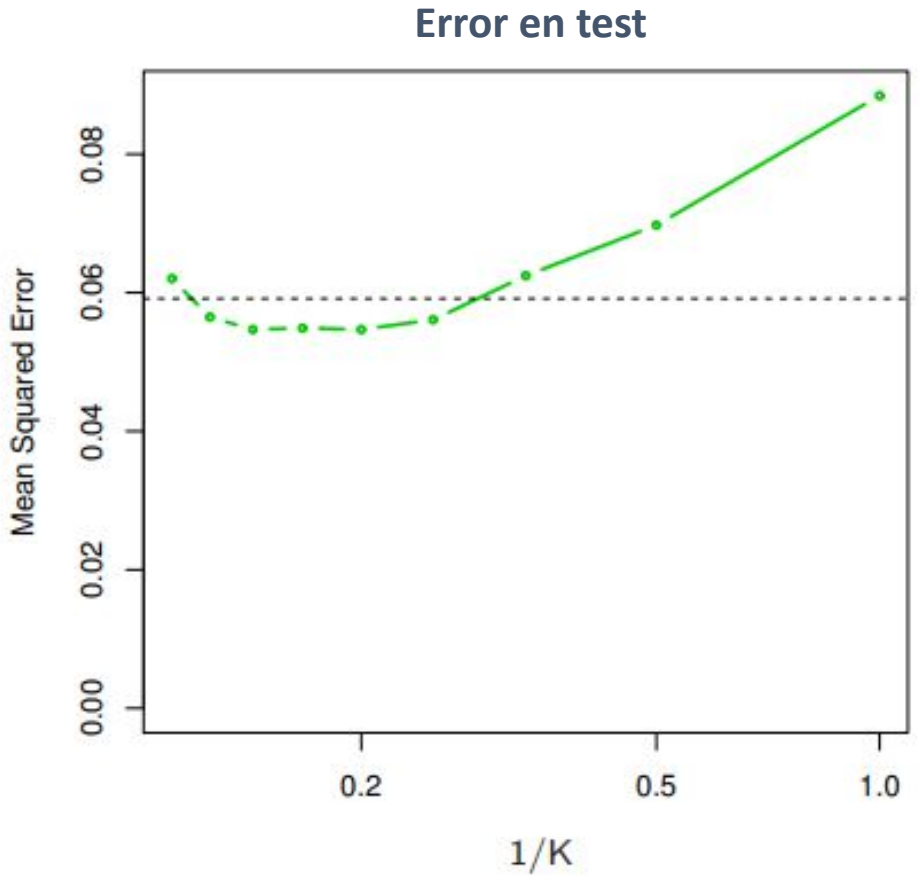
KNN

MCO

Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

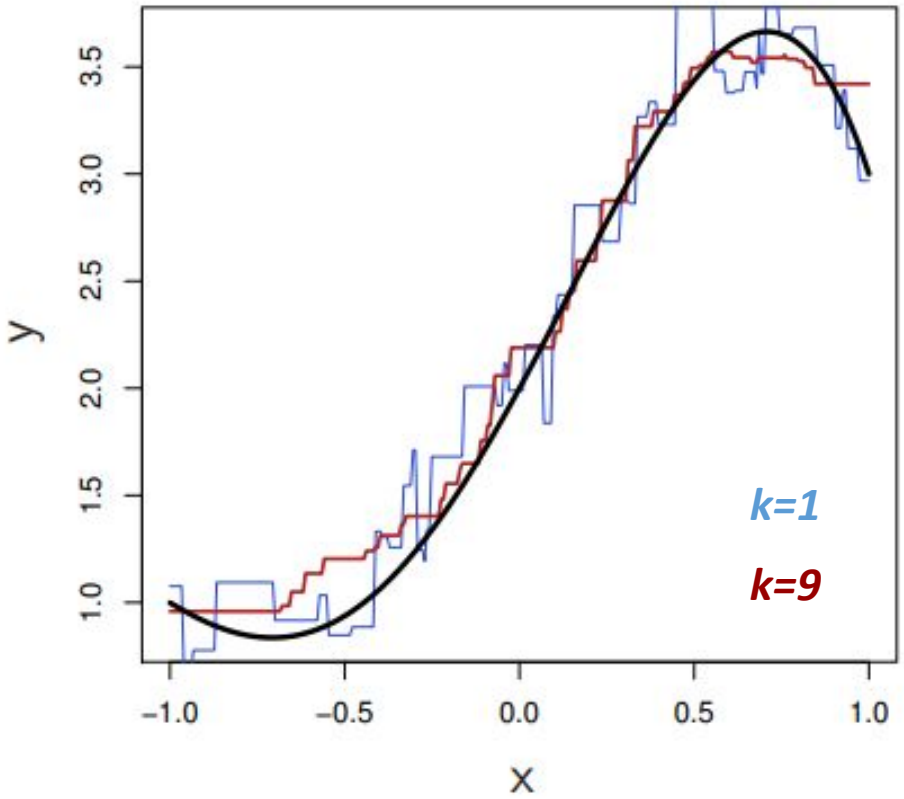
KNN regression

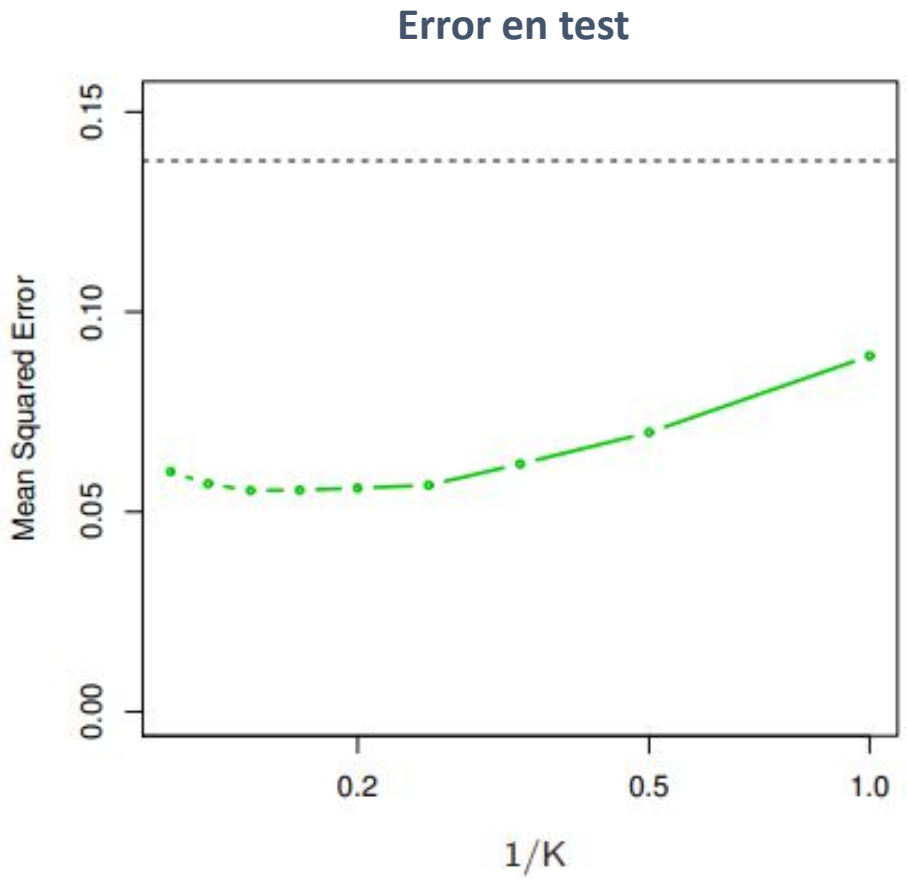




Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

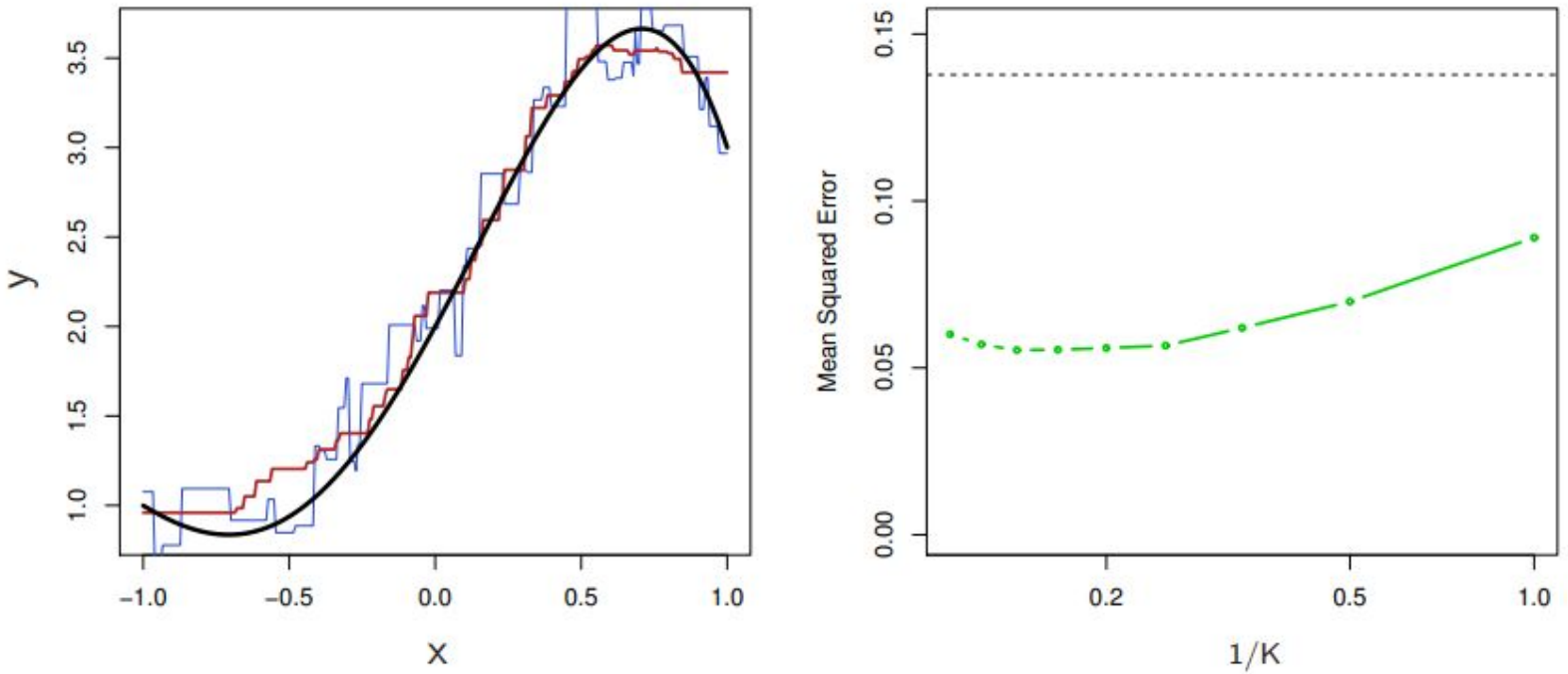
KNN regression





Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN vs Modelos lineales



En DGP **no lineales**, KNN tiene mejor rendimiento

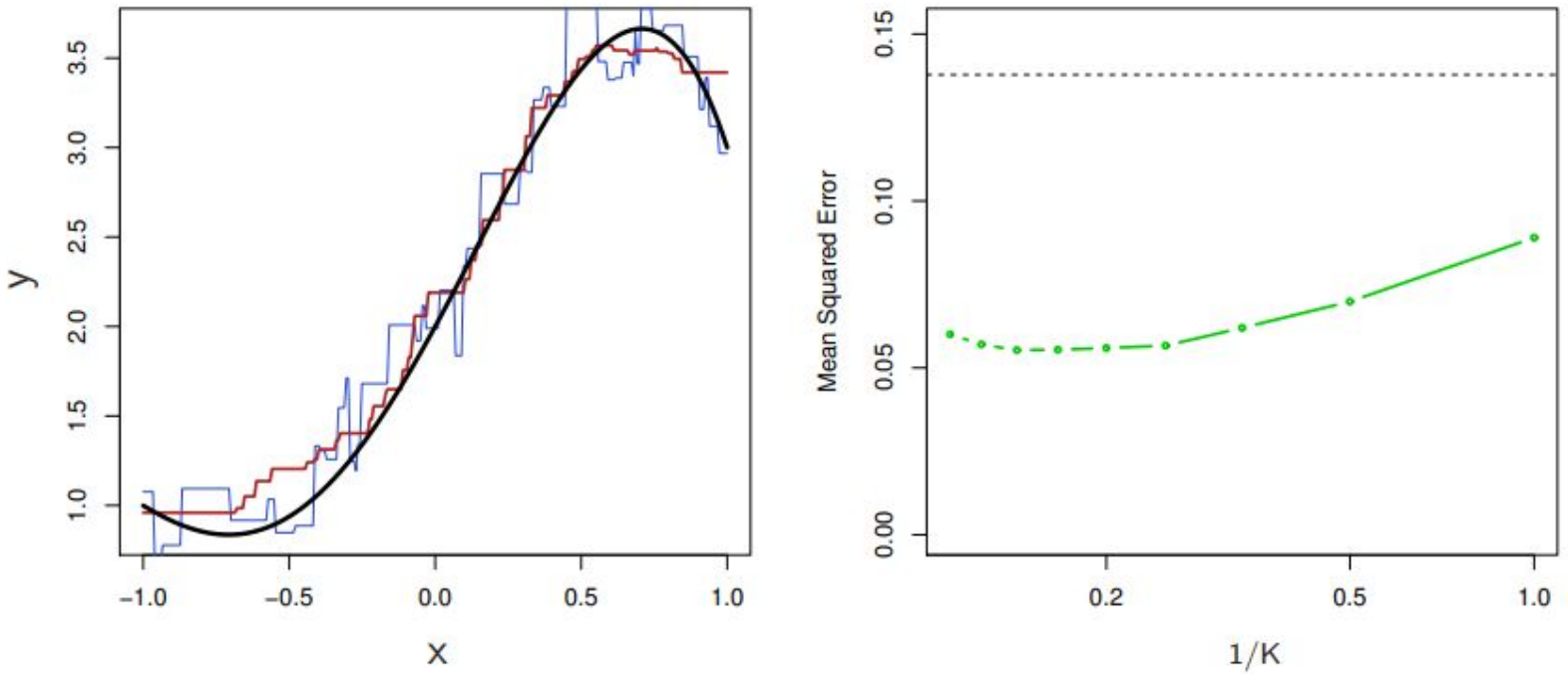
A diferencia de MCO, **es un método local y no paramétrico**

[no asume una forma específica de *f(X)*]

¿Esto quiere decir que MCO no sirve en este escenario?

Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN vs Modelos lineales

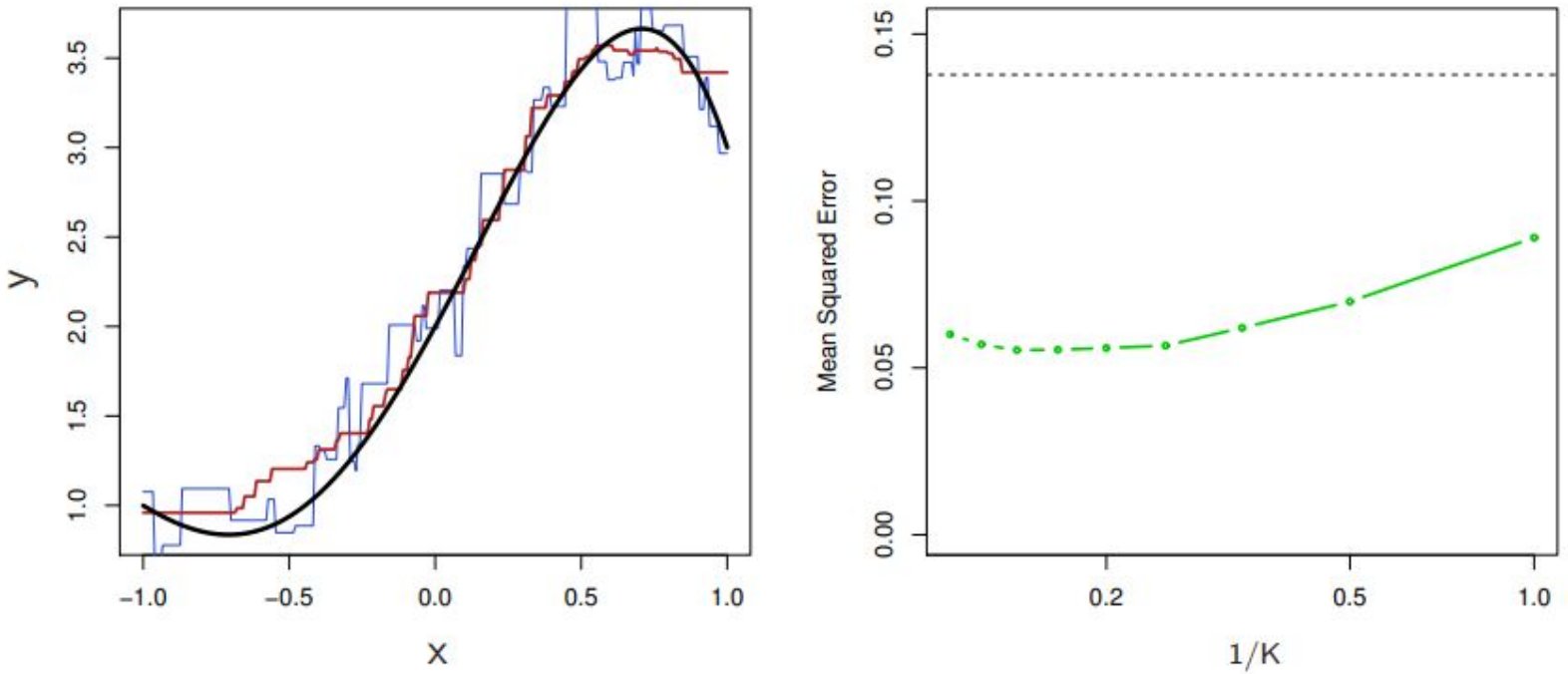


¿Esto quiere decir que MCO no sirve en este escenario?

NO! Podríamos transformar *x* para mejorar el ajuste de MCO :)

Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN vs Modelos lineales

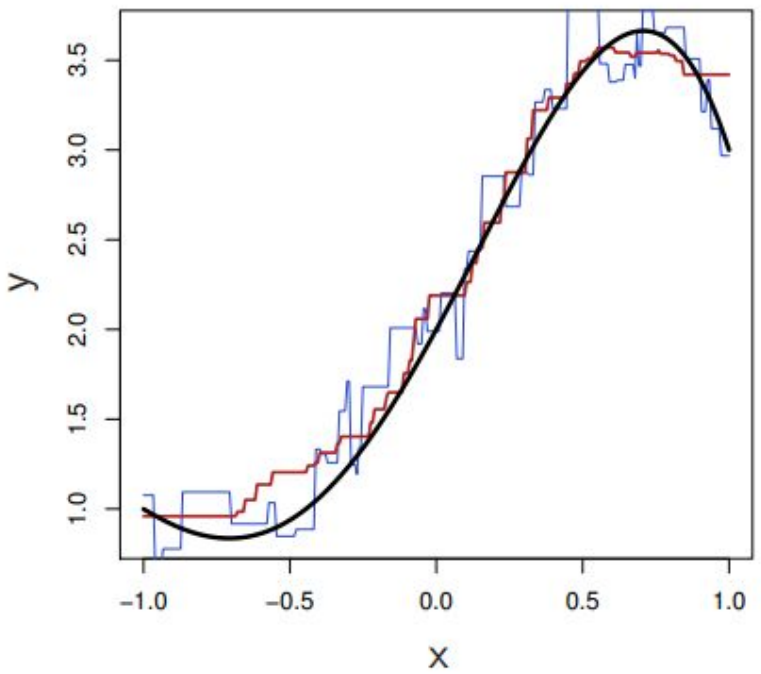


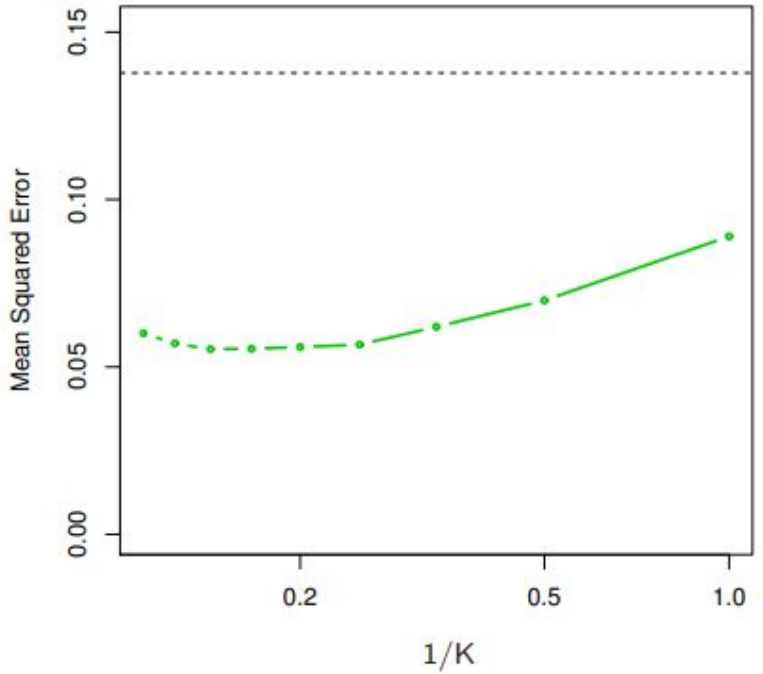
¿Cuántos *valores* necesitamos en memoria para hacer la predicción de una observación nueva con MCO?

¿Y con KNN?

Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

KNN vs Modelos lineales





¿Cuántos *valores* necesitamos en memoria para hacer la predicción de una observación nueva con MCO? β_0 y β_1

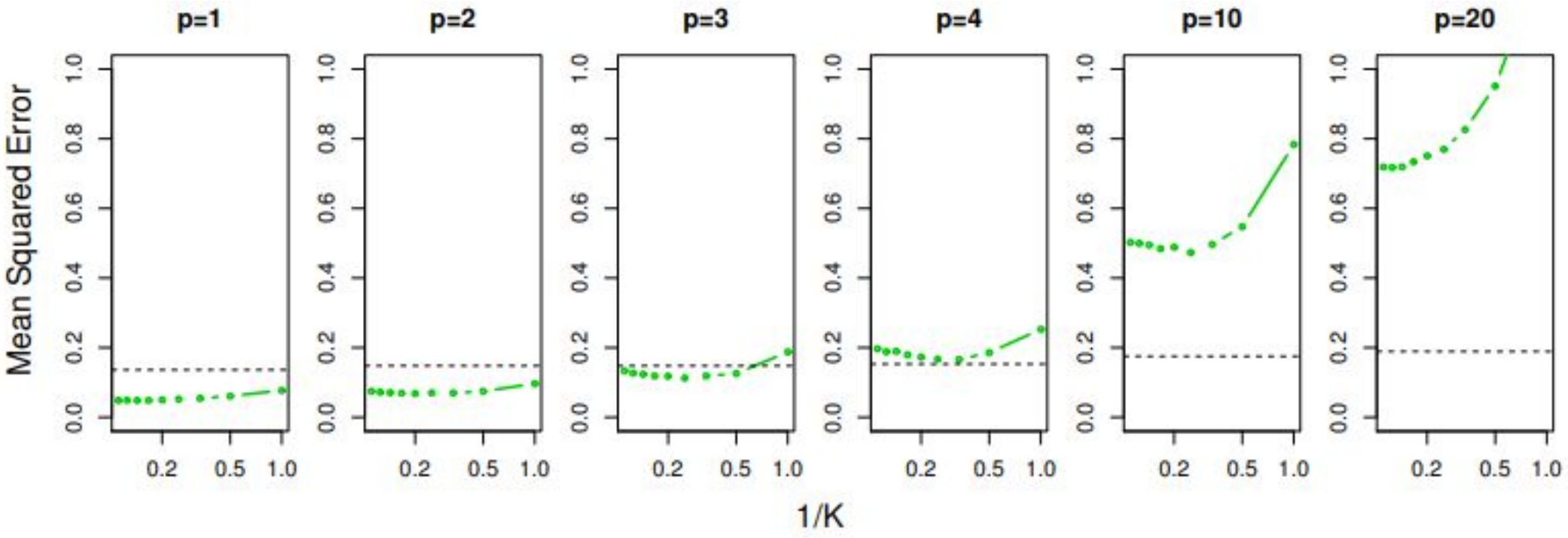
¿Y con KNN? Las matriz de observaciones *n x p*

⟶ KNN es costoso en términos de memoria y cómputo (necesitamos calcular la matriz de distancias y encontrar los mínimos)

Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

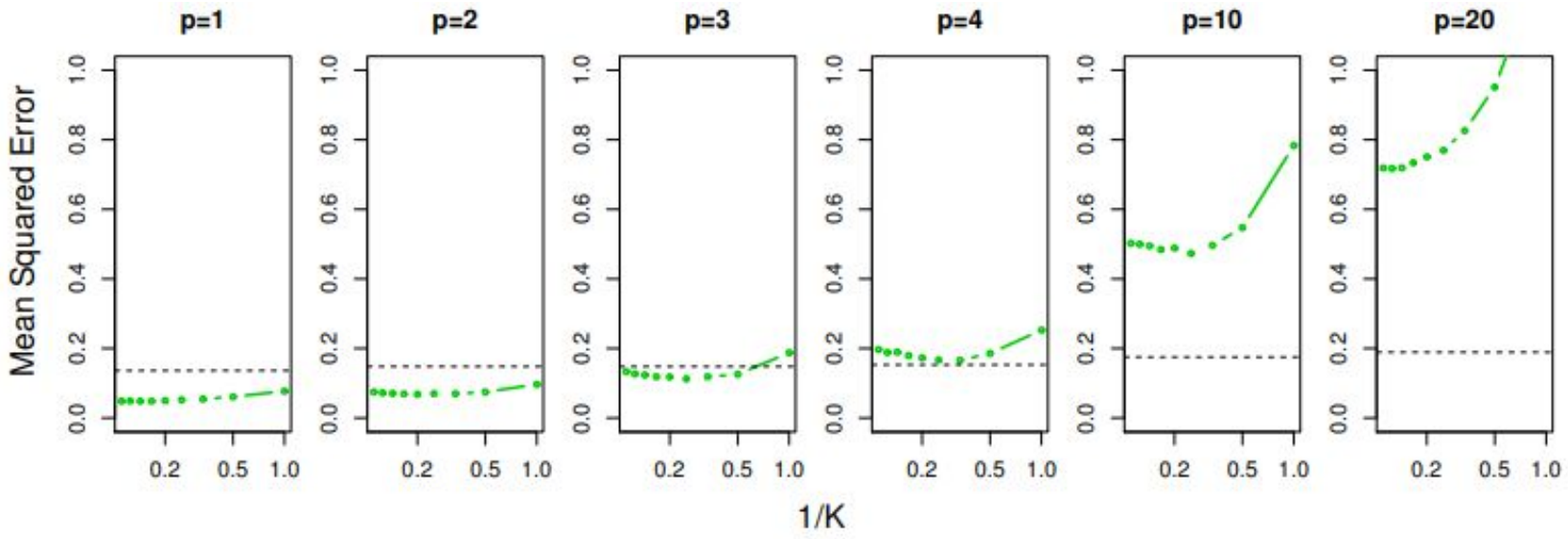
KNN vs Modelos lineales

¿Qué pasa si agregamos covariables irrelevantes al escenario anterior?
j=2,...,20



Fuente: [An Introduction to Statistical Learning (James *et al*, 2021)]

KNN y *La Maldición de la Dimensionalidad*



¡Las distancias se deterioran mucho en dimensión alta! (alto *p/N*)
→ Para seguir encontrando individuos parecidos a medida que aumentamos *p*, necesitamos aumentar *N* exponencialmente

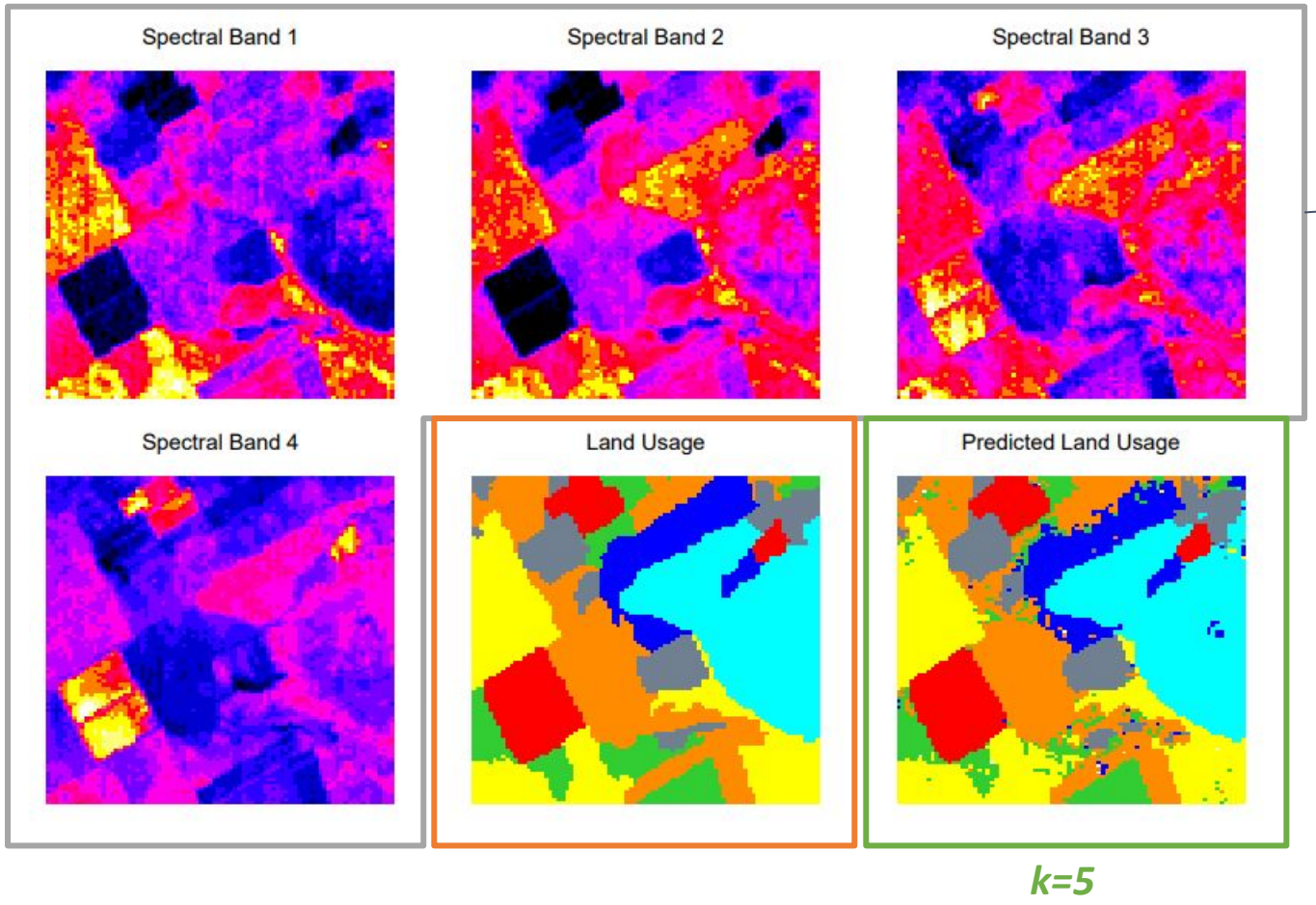
Fuente: An Introduction to Statistical Learning (James *et al*, 2021)

Aplicación en imágenes satelitales

¿Cuál es el uso de la tierra Y de cada píxel x?

(Michie et al, 1994)

82 x 100

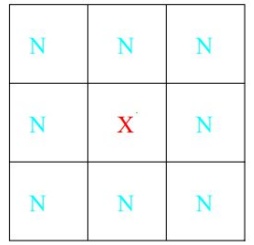


k=5

$$y \in \{\text{red soil, cotton, gray soil, mixture, ...}\}$$

$$X$$

featurización



1 pixel + 8 neighbors

$$\phi(X) \in \mathbb{R}^{36}$$

feature map

$$\hat{y} = \hat{f}(\phi(X))$$

KNN

Fuente: The Elements of Statistical Learning (Hastie *et al*, 2009)

Lecturas recomendadas

- *An Introduction to Statistical Learning*

  2.2.3 + 3.5

- *The Elements of Statistical Learning*

  2.3 + 2.9 + 13.3 + 14.3.2 + 14.3.3

- *Probabilistic Machine Learning: An Introduction*

  16.1