

KNN, K VECINOS MÁS CERCANOS

ALAN REYES-FIGUEROA
APRENDIZAJE ESTADÍSTICO

(AULA 21) 28.MARZO.2025

KNN, K vecinos más cercanos

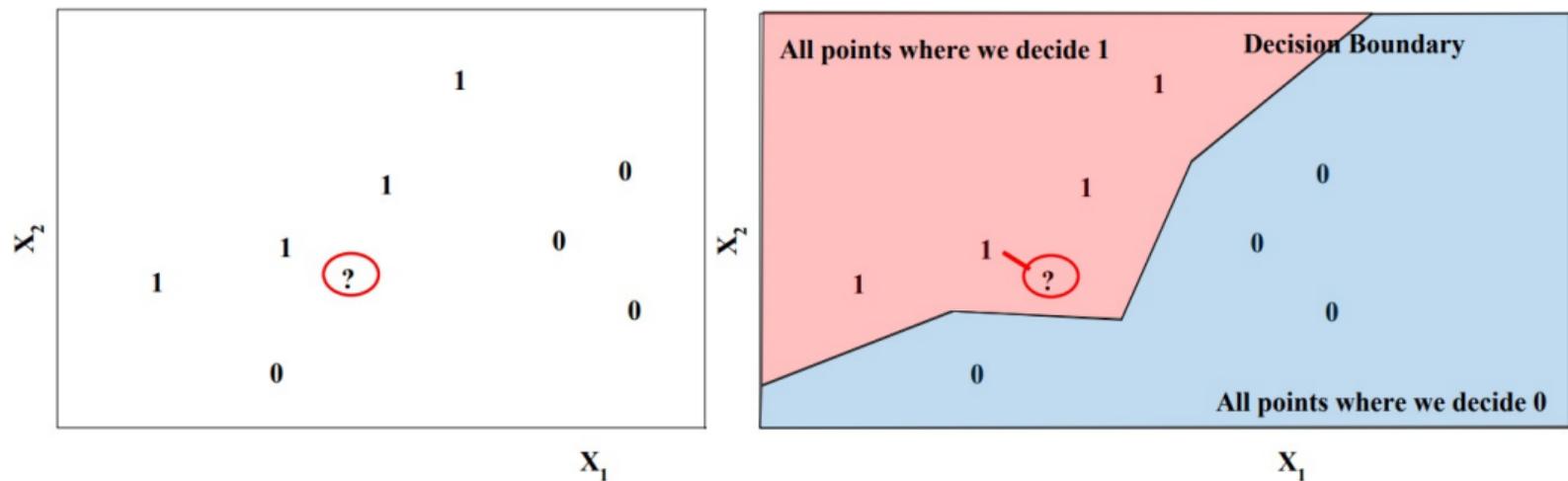
Consideramos el conjunto de datos $\{(\mathbf{x}_i, y_i)\}$, con $\mathbf{x}_i \in \mathbb{R}^d$
(en ocasiones denotamos $\mathbb{X} = (\mathbf{x}_i) \in \mathbb{R}^{n \times d}$, $Y = (y_i) \in \mathbb{R}^n$).

Dado $\mathbf{x} \in \mathbb{R}^d$, para decidir el valor de $\hat{y}(\mathbf{x})$, construimos $N_k(\mathbf{x})$ el conjunto de las k observaciones más cercanas a \mathbf{x} .

- Para clasificación: decidimos por votación, esto es, asignamos a \mathbf{x} la categoría más frecuente en $\{y_i : i \in N_k(\mathbf{x})\}$.
- Para regresión: decisión por promedio, *i.e.* asignamos a \mathbf{x} el promedio de $\{y_i : i \in N_k(\mathbf{x})\}$.

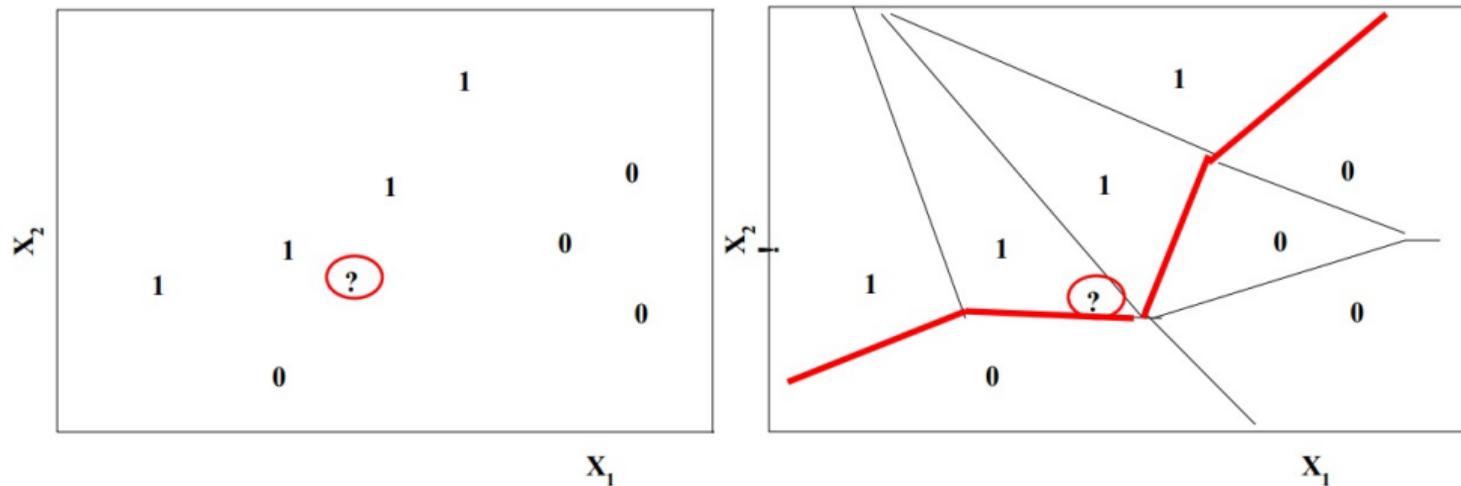
Obs! comentarios sobre cómo romper empates / métodos robustos.
El caso $k = 1$ se llama el clasificador de **vecino más cercano**.

KNN, K vecinos más cercanos



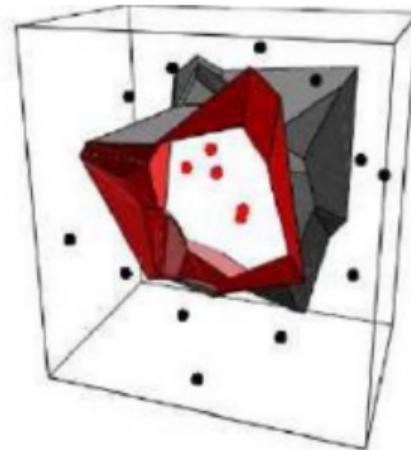
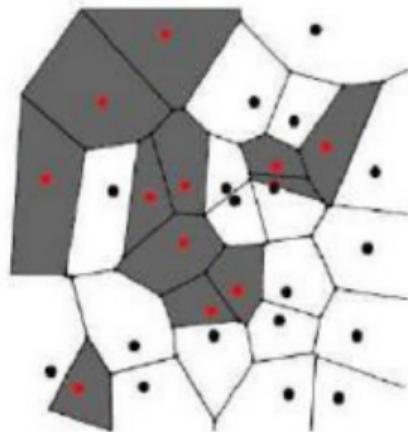
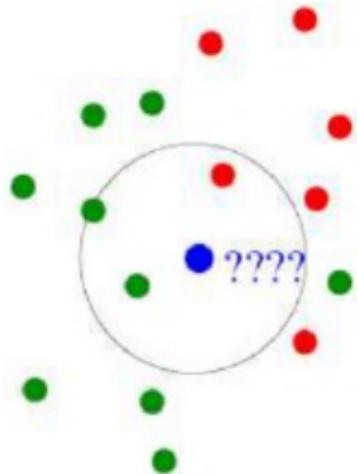
Ejemplo de KNN en el caso de clasificación.

KNN, K vecinos más cercanos



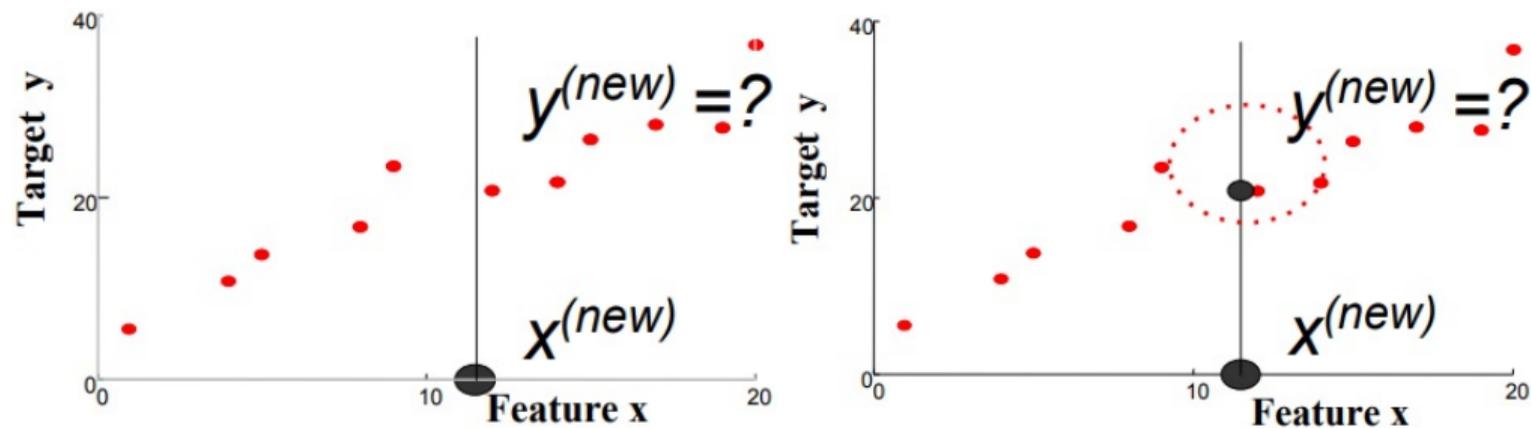
Ejemplo de KNN en el caso de clasificación. Para $k = 1$, la frontera de clasificación coincide con un diagrama de Voronoi.

KNN, K vecinos más cercanos



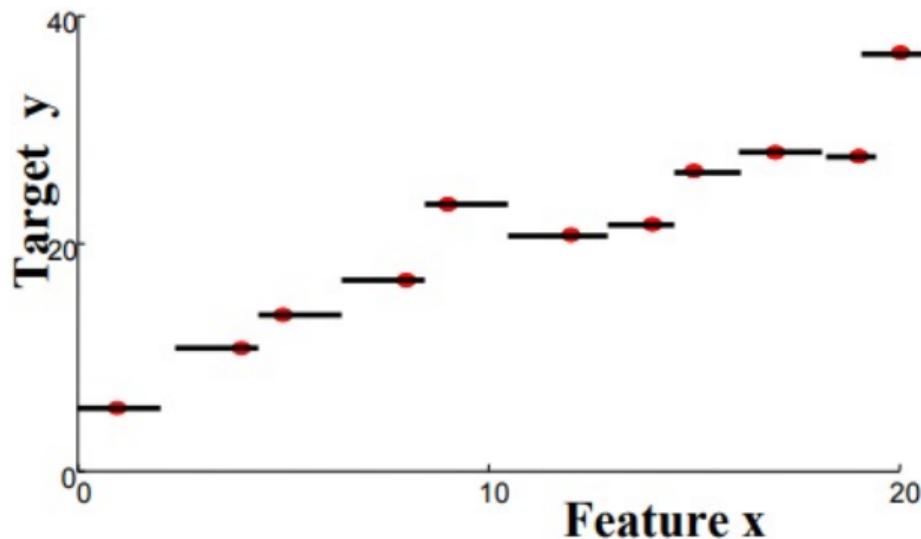
Ejemplo de KNN en el caso de clasificación. Para $k = 1$, la frontera de clasificación coincide con un diagrama de Voronoi.

KNN, K vecinos más cercanos



Ejemplo de KNN en el caso de regresión.

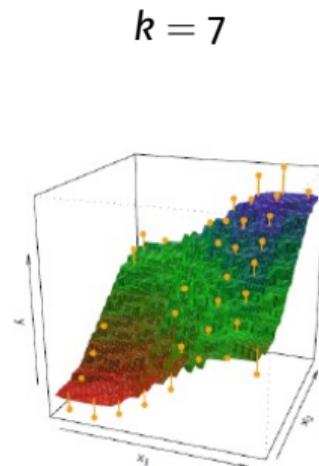
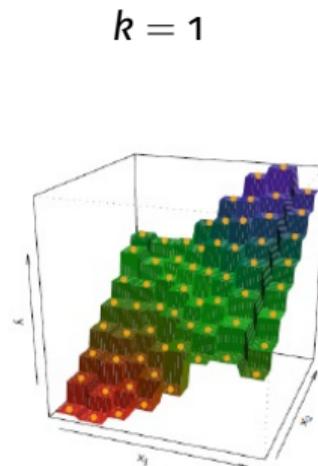
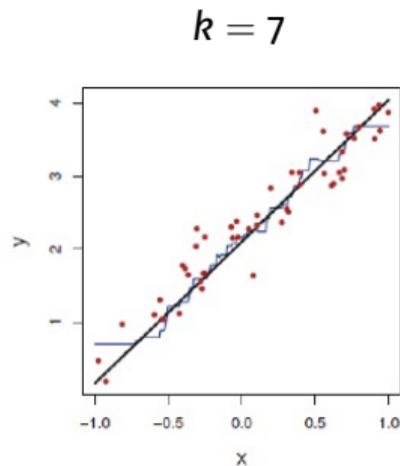
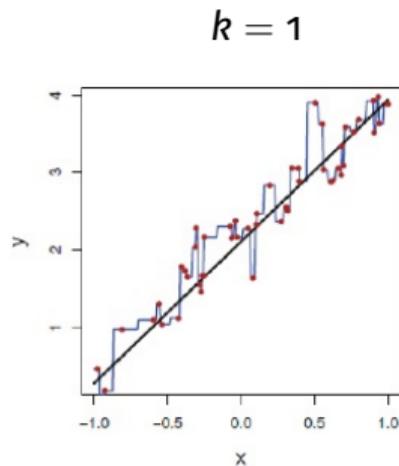
KNN, K vecinos más cercanos



Ejemplo de KNN en el caso de regresión en el caso $k = 1$. Las discontinuidades ocurren en los puntos medios entre dos observaciones consecutivas.

KNN, K vecinos más cercanos

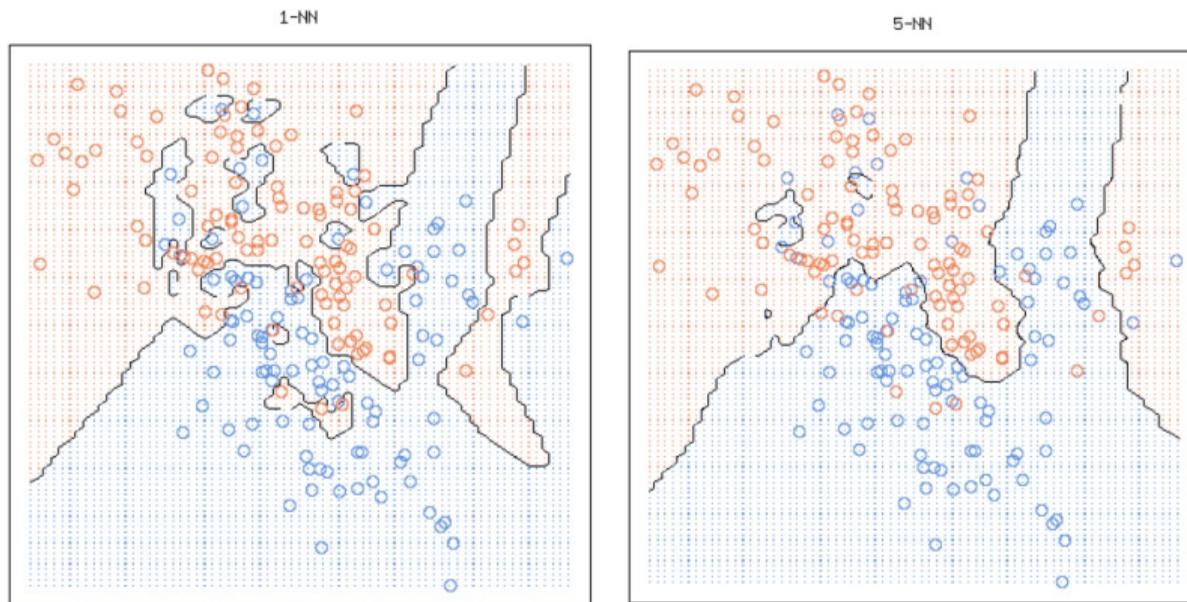
Comportamiento al variar el valor de k :



Ejemplo de *KNN* en el caso de regresión.

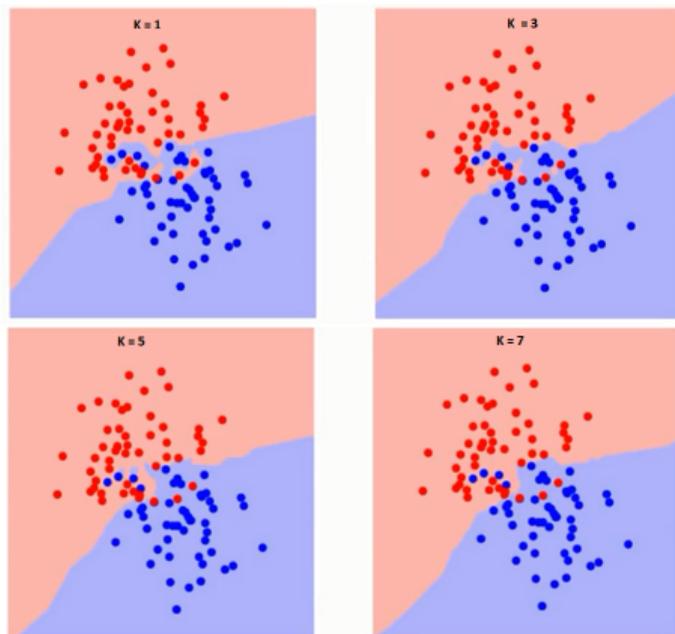
KNN, K vecinos más cercanos

Comportamiento al variar el valor de k :



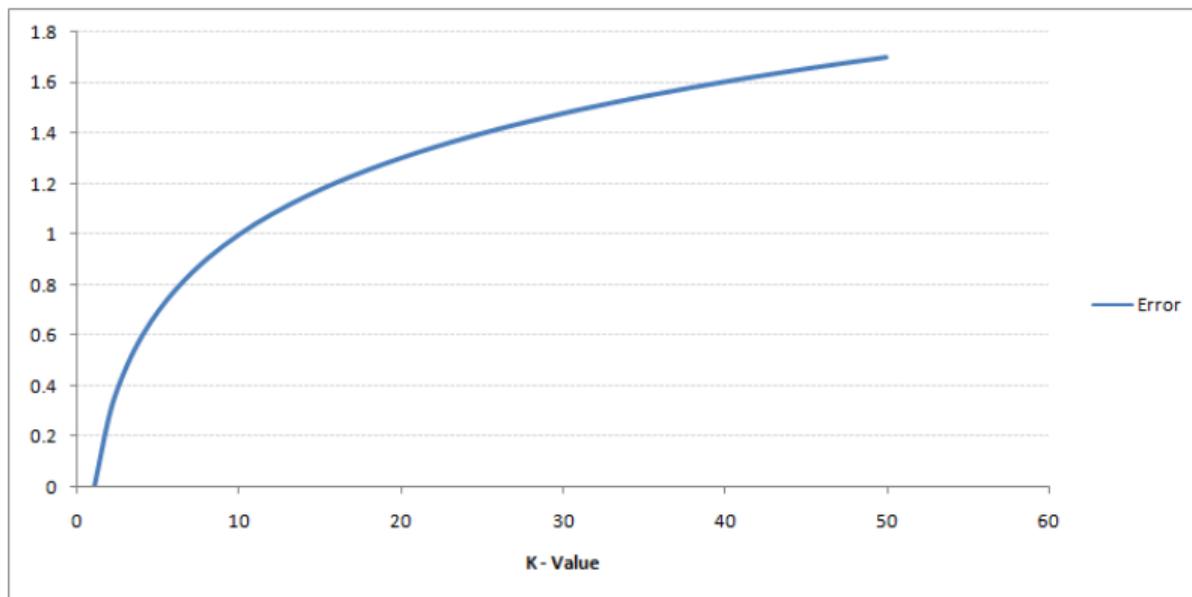
KNN, K vecinos más cercanos

Al aumentar k las fronteras de clasificación se suavizan.



KNN, K vecinos más cercanos

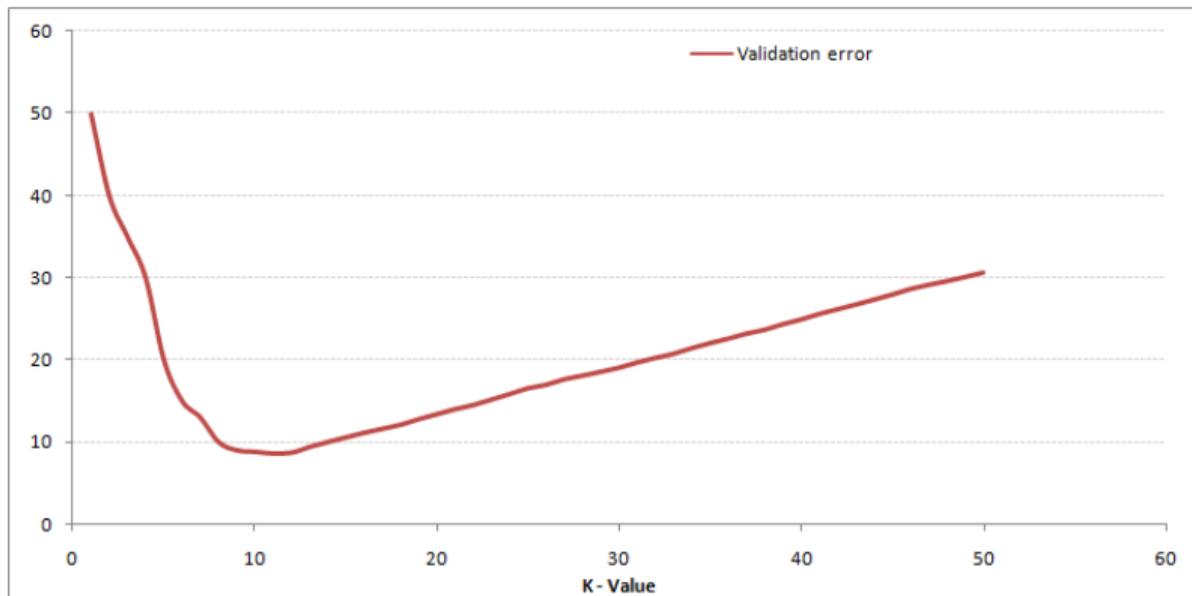
¿Cómo elegir k ?



Error de entrenamiento en KNN .

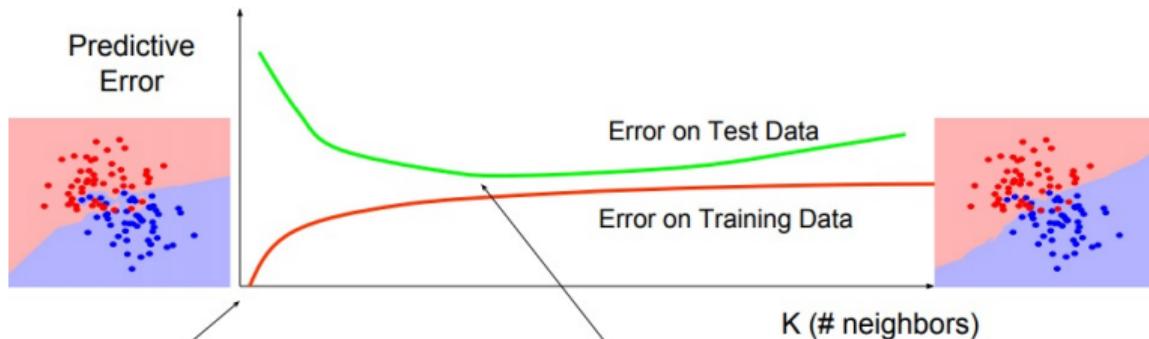
KNN, K vecinos más cercanos

¿Cómo elegir k ?



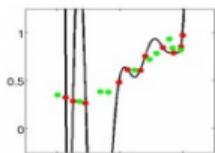
Error de validación en KNN.

KNN, K vecinos más cercanos

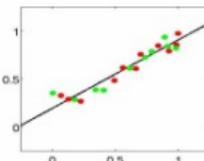
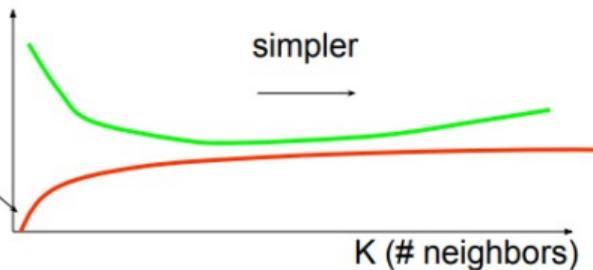


K=1? Zero error!
Training data have been memorized...

Best value of K

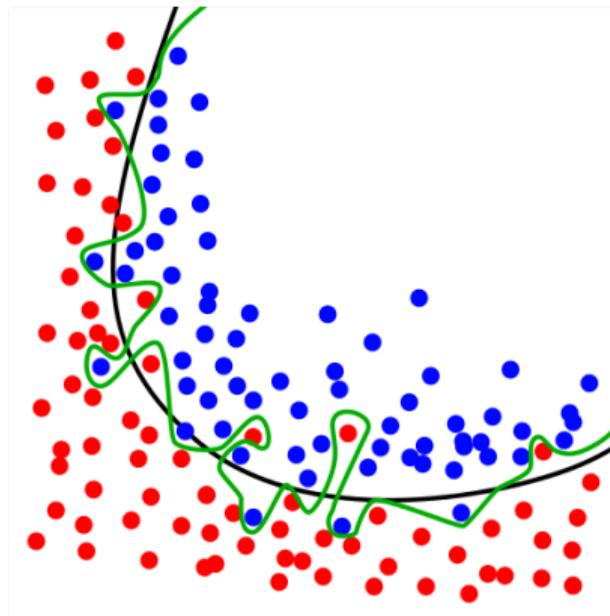


Too complex



KNN, K vecinos más cercanos

Pregunta: ¿Cómo medir la complejidad en KNN?



Está relacionada con el valor $\frac{1}{R}$: Complejidad(KNN) = $\frac{1}{R}$.