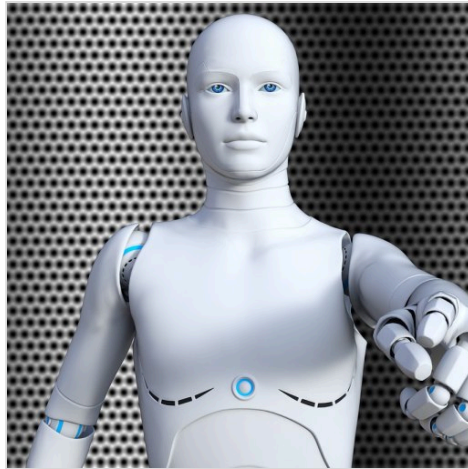


L U M B R E



Vecinos más cercanos (KNN)

Grado: 5

Materia: Aprendizaje Supervisado · Machine Learning

11/10/2026

Antes de empezar

El modelo más humilde del repertorio: «dime quiénes son tus vecinos y te diré qué eres». Los k vecinos más cercanos no aprenden ningún parámetro; guardan los datos y deciden mirando a los ejemplos más parecidos. Simple, revelador y cargado de lecciones.

QUÉ APRENDERÁS

- Clasificar y predecir por semejanza local con k vecinos más cercanos, eligiendo k y la distancia, y comprendiendo su dependencia del escalado y de la dimensionalidad.

Vecinos más cercanos

Aprender sin aprender

knn es el modelo que no entrena: memoriza todos los puntos y, ante uno nuevo, mira a sus k vecinos más próximos y vota. Su crudeza lo hace inmejorable para intuir qué es una frontera de decisión y por qué el escalado y las dimensiones importan tanto.

VIDEO

K-vecinos más cercanos aplicado con Python

Vídeo en español que explica el algoritmo k -vecinos más cercanos y lo aplica con Python.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=kqtcotJKXpA>

Un modelo perezoso

knn es un aprendiz perezoso: no generaliza hasta el momento de predecir, cuando busca en los datos guardados. Toda la «inteligencia» reside en la métrica de distancia y en el vecindario local. No hay parámetros que ajustar más allá de elegir k .

Cómo clasifica

Para un nuevo punto, encuentra los k ejemplos entrenados más cercanos y asigna la clase mayoritaria entre ellos. En regresión, devuelve el promedio de esos vecinos. Es un votante local: la respuesta depende solo de quién está al lado.

La elección de k

k pequeño: el modelo sigue el ruido, fronteras rugosas, alta varianza, sensible a outliers. k grande: suaviza y sesga, hasta que con k igual al total siempre predice la clase mayoritaria. K impar en binaria evita empates. Es el único hiperparámetro y manda.

Todo depende de la distancia

La distancia euclídea mide vecindad en las unidades originales. Si una característica varía entre cero y un millón y otra entre cero y uno, la primera domina y la segunda desaparece. El vecindario resultante es un artefacto de las escalas, no del mundo.

Escalar siempre

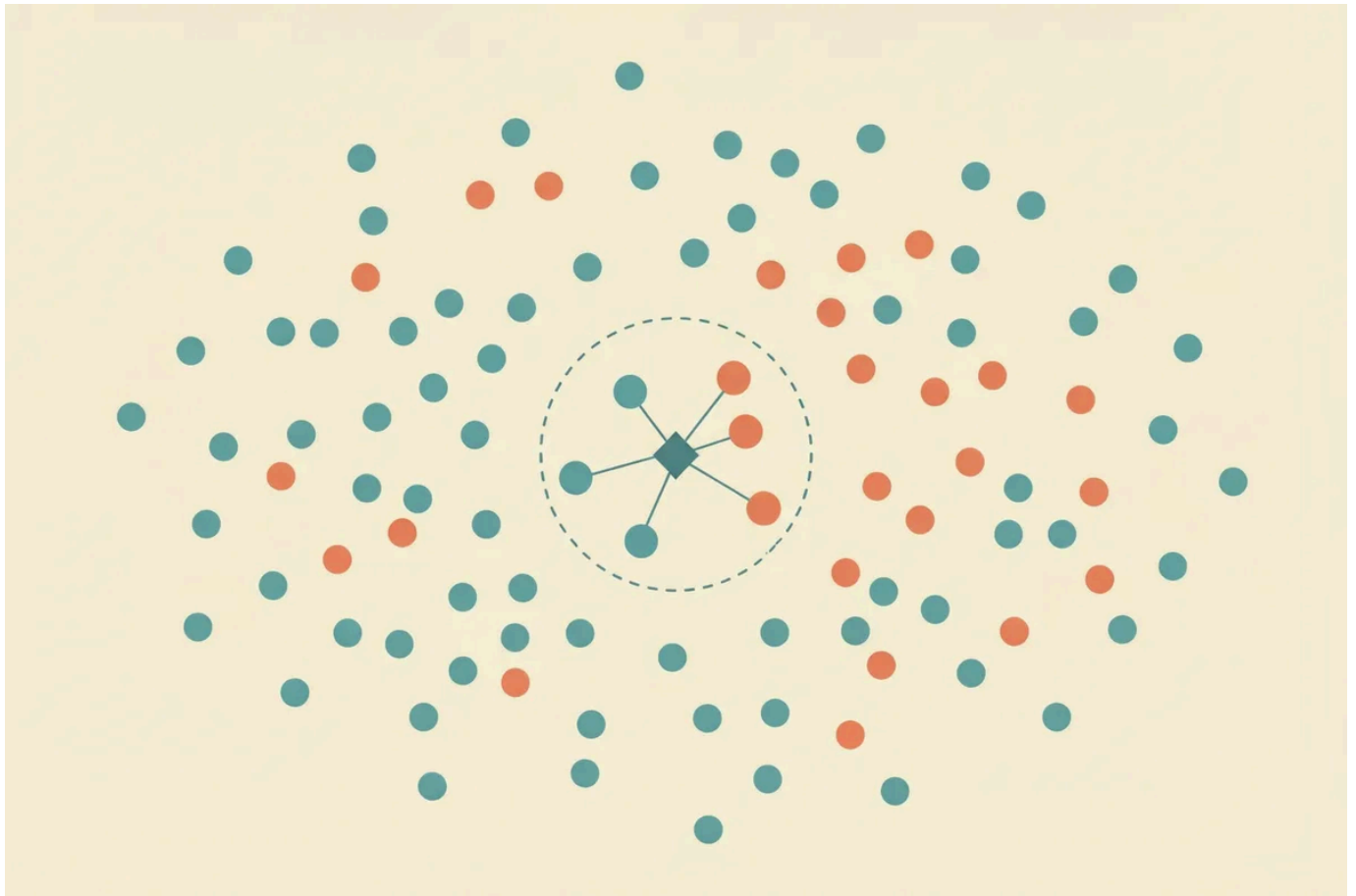
Por eso knn exige tipificar las variables a escala comparable antes de nada. Sin estandarizar, el modelo es ciego a las columnas de rango pequeño. Es el ejemplo más claro de por qué el preprocesado no es opcional.

La maldición dimensional

Al acumular dimensiones, todos los puntos acaban estando lejos de todos: las distancias se igualan y el «vecino más cercano» deja de ser cercano. knn se degrada rápido con muchas características, revelando crudamente la maldición de la dimensionalidad.

Fronteras de decisión

Su límite entre clases es local y flexible: puede dibujar formas arbitrarias siguiendo los datos. Por eso se parece a un método de ventana: mucho detalle con k pequeño, decisiones casi globales con k grande.



La predicción por vecindario: el punto nuevo no consulta ninguna fórmula, solo mira a sus k vecinos más próximos dentro del anillo y deja que voten.

Coste de predecir

Al no comprimir nada en un modelo, knn debe comparar el punto nuevo contra todo el entrenamiento en cada predicción. Barato de «entrenar», carísimo de predecir y de almacenar cuando los datos son millonarios.

Errores frecuentes

- Omitir el estandarizado y dejar que domine la variable de mayor rango.
- Usar knn con cientos de dimensiones sin reducción previa.

Ejemplo resuelto: clasificar una flor con vecinos

Se quieren asignar especies a partir de medidas de pétalos, la tarea clásica. El procedimiento:

1. Escalar: tipificar pétalos y sépalos a escala comparable para que nadie domine la distancia.
2. Medir: calcular la distancia del punto nuevo a todos los ejemplos guardados.
3. Votar: quedarse con los k más próximos y asignar la clase mayoritaria entre ellos.

4. Sintonizar k: explorar de uno en uno hasta achicar la brecha entre entrenamiento y validación.
5. Desconfiar en alto: si hay muchas columnas, reducir dimensiones o elegir otro modelo.

¿Para qué sirve en la realidad?

El vecindario local explica recomendadores de «clientes como tú», sistemas de búsqueda por imagen y detección de anomalías por distancia a lo conocido. Además, knn es el mejor maestro: quien lo entiende entiende escalado, dimensión y el compromiso sesgo-varianza de una vez.

1. En k vecinos más cercanos, disminuir el valor de k tiende a:

- ☐ Aumentar el sesgo y reducir la varianza
- ☐ Reducir el sesgo pero aumentar la varianza y la sensibilidad al ruido
- ☐ No cambiar nada
- ☐ Hacer el modelo más rápido de predecir

2. knn requiere estandarizar las características antes de calcular distancias.

- ☐ Verdadero
- ☐ Falso

Por qué enseñar knn primero

Aunque rara vez gana, knn cristaliza ideas de todo el campo: vecindad, escalado, la maldición dimensional y el sesgo-varianza en un solo dial. Quien entiende por qué knn falla en alta dimensión entiende algo profundo sobre casi todos los modelos.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. KNN



4. Ordena la predicción con k vecinos más cercanos:

Elementos:

- tipificar las características
- medir la distancia al nuevo punto
- quedarse con los k más próximos
- votar la clase mayoritaria o promediar

Escalar	. . .
Distancia	. . .
Vecinos	. . .
Voto	. . .

Explained <https://explained.ai/Visualizaciones de fronteras de decisión y del papel de k en vecinos cercanos>.

5. Une cada característica de knn con su consecuencia práctica.

5.	no entrena, guarda los ejemplos	-----
5.	mide distancias en las unidades originales	-----
5.	k pequeño sigue el ruido	-----
5.	las distancias se igualan al crecer dimensiones	-----

Opciones: predecir es caro con millones de datos · hay que escalar siempre · alta varianza y fronteras rugosas · maldición de la dimensionalidad

Aprender sin entrenar

K-vecinos más cercanos

Idea

voto local

sin parámetros

Decisiones

elegir k

elegir distancia

Cuidados

escalar

reducir dimensiones

Costes

guardar todo

predecir despacio

El mapa mental interactivo es interactivo en la versión online.



¿Qué es el algoritmo de k vecinos más cercanos? <https://www.ibm.com/es-es/think/topics/knn> Explicación de referencia en español sobre knn y sus usos.

6. Con k igual a uno tu modelo clava el entrenamiento pero falla en validación. ¿Qué tipo de error es y cómo lo confirmarías variando k?

Referencia curricular

Grado:	5
Materia:	Aprendizaje Supervisado · Machine Learning

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-supra-001-vecinos-mas-cercanos-knn>

Material educativo generado desde Lumbre · lumbre.lat