

**Centro Universitario de Ciencias Exactas e Ingenierías**

**Ingeniería Informática**

**Sistemas basados en conocimientos**

**D01 – 2024A**

**Algoritmo kNN**

**Eduardo Rangel Heras**

**Cameros Moran Kevin Eduardo - 220793406**

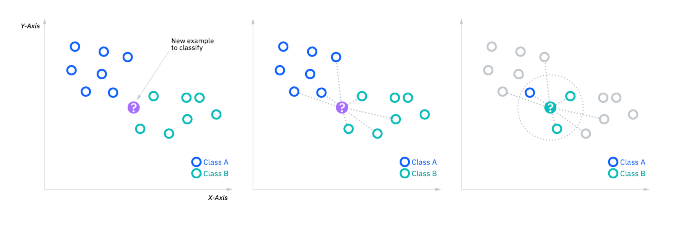
**Algoritmo kNN**

¿Qué es KNN?

El **algoritmo de k vecinos más cercanos**, también conocido como **KNN o k-NN**, es un clasificador de aprendizaje supervisado no paramétrico, que utiliza la proximidad para hacer clasificaciones o predicciones sobre la agrupación de un punto de datos individual.

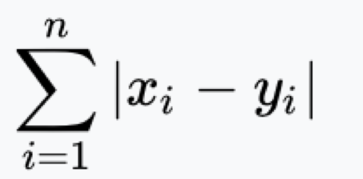
Si bien se puede usar para problemas de **regresión o clasificación**, generalmente se usa como un algoritmo de clasificación, partiendo de la suposición de que se pueden encontrar puntos similares cerca uno del otro.

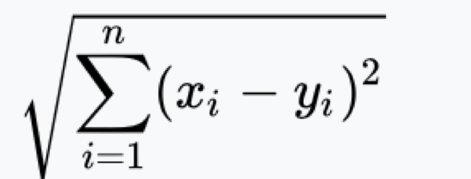
Para los problemas de clasificación, se asigna una etiqueta de clase sobre la base de un voto mayoritario, es decir, se utiliza la etiqueta que se representa con más frecuencia alrededor de un punto de datos determinado. Si bien esto técnicamente se considera "voto por mayoría", el término "voto por mayoría" se usa más comúnmente en la literatura. La distinción entre estas terminologías es que "voto mayoritario" técnicamente requiere una mayoría superior al 50 %, lo que funciona principalmente cuando solo hay dos categorías. Cuando tiene varias clases, por ejemplo, cuatro categorías, no necesita necesariamente el 50 % de los votos para llegar a una conclusión sobre una clase; puede asignar una etiqueta de clase con un voto superior al 25 %.

****

La lógica detrás del algoritmo de las K vecinas más cercanas es una de las más sencillas de todos los algoritmos de Machine Learning supervisados:

* **Etapa 1:** Seleccionar el número de K vecinas
* **Etapa 2:** Calcular la distancia

Euclidienne

Manhattan

Desde un punto no clasificado a otros puntos:

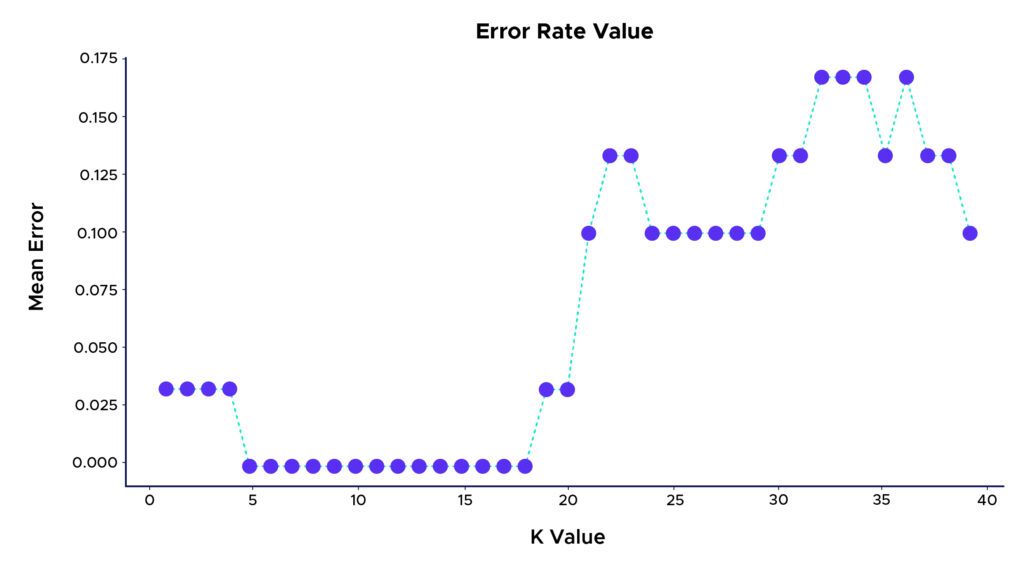
* **Etapa 3:** tomar las K vecinas más cercanas según la distancia calculada
* **Etapa 4:** entre las K vecinas, contar el número de puntos en cada categoría.
* **Etapa 5:** atribuir un nuevo punto a la categoría más presente entre las K vecinas
* **Etapa 6:** El modelo está listo.

**KNN : un ejemplo sobre su uso.**

A partir de ahora podemos analizar un ejemplo sobre el uso del algoritmo de las K vecinas. Gracias a la biblioteca **Scikit-Learn**, podemos importar la función Neighbors Classifier que utilizaremos sobre un conjunto de datos IRIS.



En el ejemplo, gracias al algoritmo KNN, obtenemos una excelente tasa de clasificación correcta de las plantas cercana al 100%. Podemos también considerar una forma de elegir la**K que hace que la clasificación sea aún mejor**. Una forma de encontrarla consiste en trazar el gráfico del valor K y la tasa de error correspondiente al conjunto de datos :



**Conclusión:**

El algoritmo kNN sigue siendo una herramienta valiosa en el repertorio de técnicas de aprendizaje automático. Su simplicidad y eficacia en diversos contextos lo hacen ideal para una variedad de aplicaciones. No obstante, su uso eficiente requiere una cuidadosa consideración de sus limitaciones y la implementación de mejoras cuando sea necesario. La continua investigación y desarrollo de técnicas avanzadas para la selección de características, reducción de dimensionalidad y métodos de búsqueda eficiente son esenciales para maximizar el potencial de kNN en problemas complejos y de gran escala.