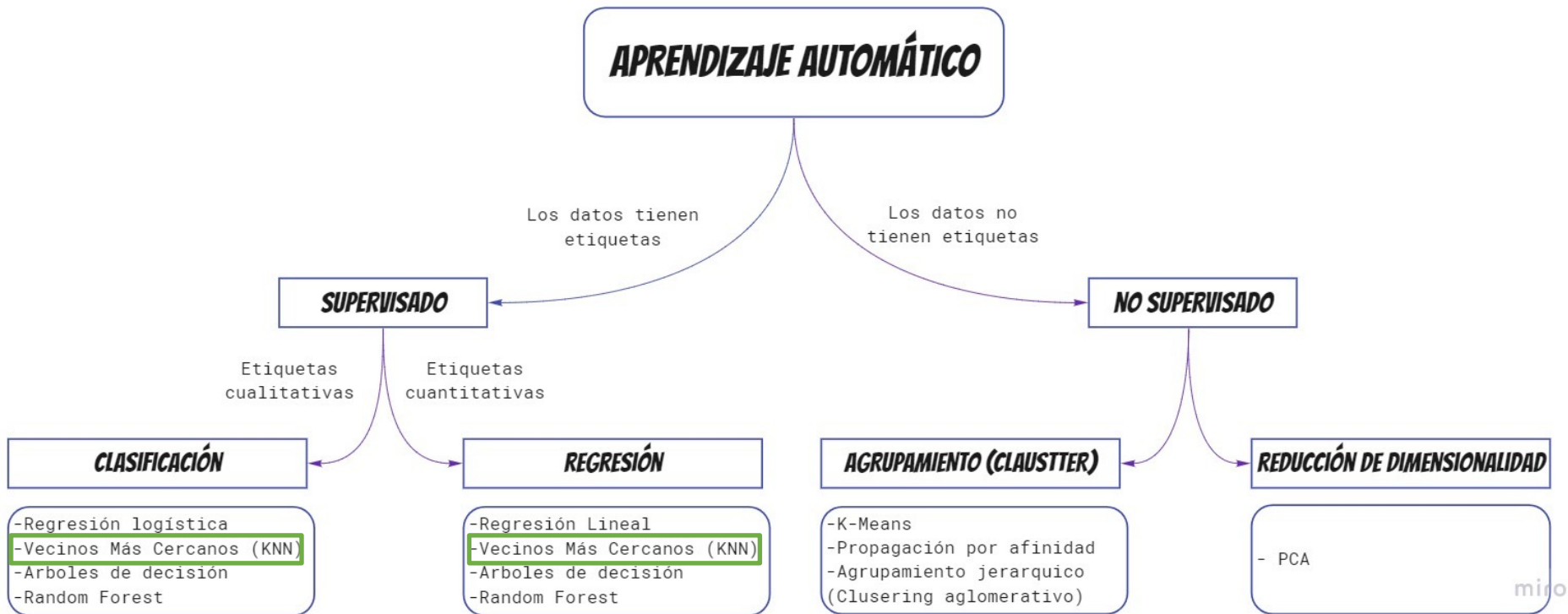


Ciencia de Datos

- Módulo 3

Overfitting y Underfitting

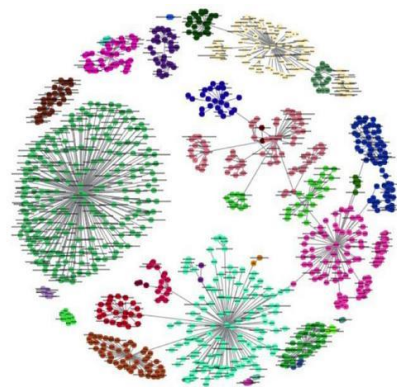




KNN – K Vecinos más cercanos

K Vecinos más Cercanos, o mejor conocido como KNN (_K Nearest Neighbors_) es un algoritmo de aprendizaje supervisado _lazy_ (perezoso) no paramétrico.

Cuando se dice que una técnica no es paramétrica, significa que no hace suposiciones sobre la distribución de datos subyacente, como en el caso de las regresiones (que asumen que los datos siguen la distribución de las regresiones). Esto es bastante útil, ya que en el mundo real, la mayoría de los datos prácticos no obedecen a los supuestos teóricos típicos que se hacen.



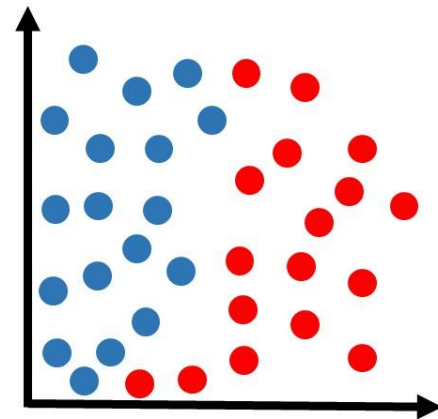
Fuente:
<https://programmerclick.com/article/85601071402/>



KNN

Es un algoritmo lazy. Lo que esto significa es que no utiliza los puntos de datos de entrenamiento para hacer ninguna generalización. En otras palabras, no hay una fase de formación explícita o es muy mínima.

La mayoría de los algoritmos lazy, especialmente KNN, toman decisiones basadas en todo el conjunto de datos de entrenamiento. Esto significa que la fase de entrenamiento es bastante rápida: hay un fase de capacitación mínima o inexistente, pero una fase de prueba costosa. El costo es en términos de tiempo y memoria. Se necesita más memoria ya que necesitamos almacenar todos los datos de entrenamiento al mismo tiempo.



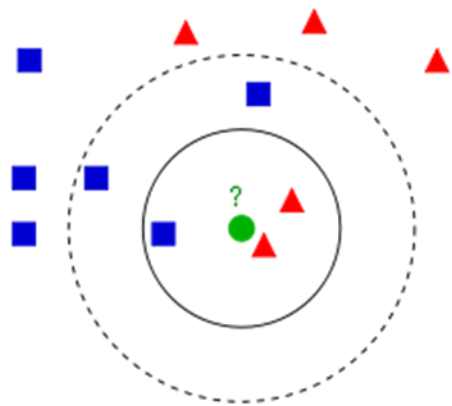
Fuente: <https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-knn/>



KNN

Como dice el refrán, “dime con quien andas y te diré quien eres”. Para hacer una predicción, KNN solo necesita que se defina un valor de K el cual representa la cantidad de “vecinos” al dato a estudiar. Esto se refiere a los K valores más cercanos a nuestro valor, dentro del set de datos. Una vez que se reciba un dato al que se le quiera predecir, sea una categoría, o valor numérico, se toman los K vecinos más cercanos y:

- * En caso de ser un problema de clasificación, se escoge la categoría más popular entre los K vecinos seleccionados.
- * En caso de ser un problema de regresión, se promedian los valores de la variable objetivo de los vecinos seleccionados.



Fuente: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm



KNN

Visualizar algoritmo

- <https://observablehq.com/@jwilber/live-k-nearest-neighbors-classification>
- <https://observablehq.com/@antoinebrl/knn-visualizing-the-variance>

Documentación Scikit-Learn

- <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html#sklearn.neighbors.KNeighborsRegressor>
- <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier>

Adicional

- <https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>
- <https://www.youtube.com/watch?v=zBCcAtg3P4k>

