

Práctica No. 8

Algoritmo K-NN

Implementación del algoritmo K-NN

El algoritmo K-NN se implementa de la siguiente manera.

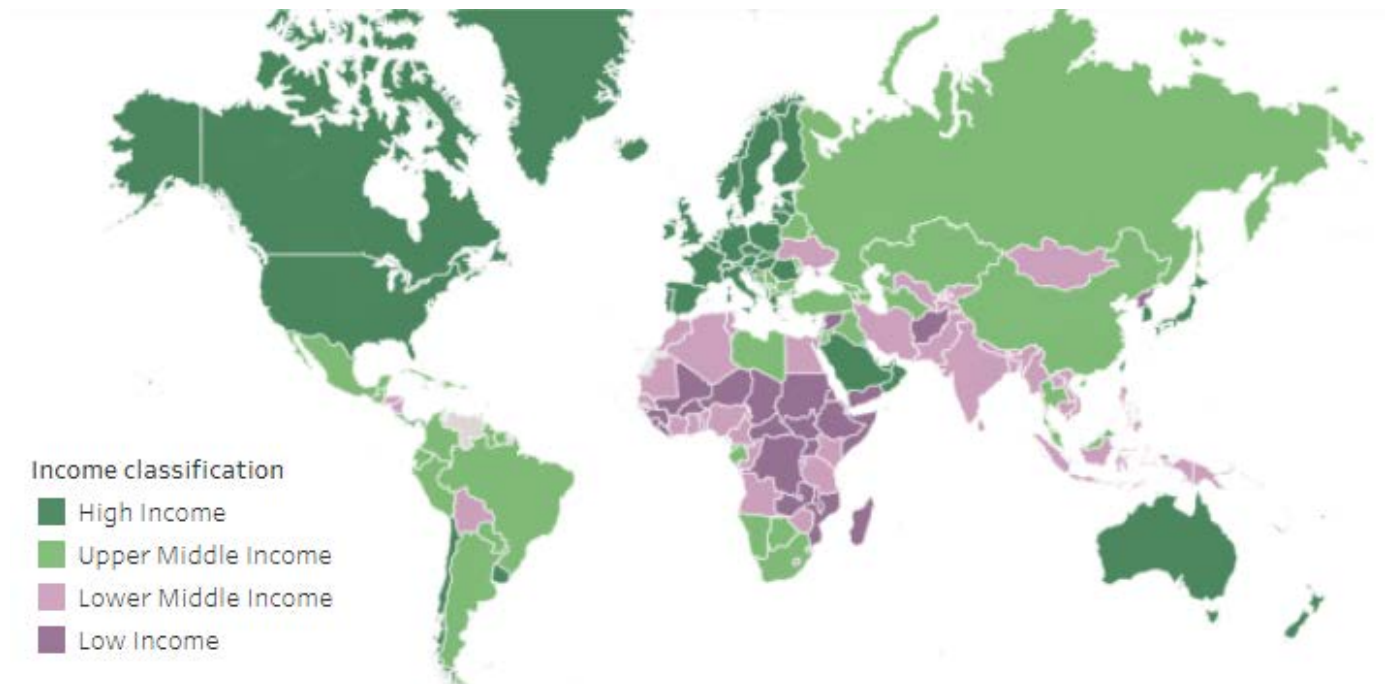
- Preparar el conjunto etiquetado de datos (se conoce la clase de todas las instancias del conjunto)
- Obtener los valores de los atributos de la instancia cuya clase se va a predecir
- Utilizando los valores del paso anterior calcular las distancias entre la instancia y el resto de las instancias del conjunto de datos
- Ordenar de menor a mayor las instancias del conjunto de datos con base en las distancias
- Seleccionar las primeras K instancias del conjunto de datos, estas instancias serán la vecindad de la instancia de clase desconocida
- Predecir la clase de la instancia de clase desconocida que será la más votada entre los K vecinos

El algoritmo K-NN pertenece a la categoría de los algoritmos denominados *lazy learning* (*aprendizaje perezoso*). El *aprendizaje perezoso* se refiere a procesos de aprendizaje de máquina en los que la generalización de los datos de entrenamiento se retrasa hasta que se hace una consulta para realizar una predicción, a pesar de su nombre, los clasificadores perezosos son muy útiles cuando se trabaja con grandes conjuntos de datos que se actualizan de frecuentemente o en tiempo real y no se puede estar entrenando un algoritmo cada vez que se agregan instancias o cambian los valores de los atributos.



Ejercicio

Considerando el siguiente mapa (<https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2022-2023>), elija 100 países para su conjunto de entrenamiento, procurando que las cuatro clases ('High', 'Upper Middle', 'Lower Middle' y 'Low'), estén representadas equitativamente, aparte elija 20 países para su conjunto de pruebas.



Con los países elegidos, construya las tablas de datos como sigue:

Pais	Esperanza de vida al nacer (Life expectancy at birth)	PIB per cápita \$US (GDP per cápita)	Emisiones de CO ₂ per cápita (CO ₂ emissions)	Clase
México	70	10,045.7	3.6	Upper_Middle

Los datos se encuentran en: <https://data.worldbank.org/country>

Con el conjunto de datos recabados (esperanza de vida, PIB y emisiones de CO₂), implemente y pruebe el algoritmo k-NN (realice predicciones con todo el conjunto de pruebas con diferentes valores de K).

Elabore su informe que incluya:

- Introducción
- Desarrollo (incluya la gráfica dispersión)
- Discusión de resultados
- Conclusiones

Suba su informe a la plataforma en formato PDF.