

# Tarea 1

## IMT2112 - Algoritmos Paralelos en Computación Científica

Elwin van 't Wout

13 de agosto de 2019

### Calendario

El método de “ $k$  vecinos más cercanos” es un algoritmo de clasificación supervisado comúnmente utilizado en el mundo de *big data*. En una base de datos con registros que ya cuenten con una etiqueta, un nuevo registro está etiquetada según la etiqueta más comun entre los  $k$  registros más cercanos.

### Actividades

Para la tarea, responden las preguntas y hagan las actividades siguientes.

1. Busca literatura para entender el algoritmo de “ $k$  vecinos más cercanos.”
2. Implementa el algoritmo en Python, con las condiciones siguientes.
  - a) Se puede sintetizar datos con un generador de números aleatorios.
  - b) El número de distintas etiquetas debe ser a menos tres.
  - c) La dimensión de la base de datos debe ser al menos dos y el tamaño al menos mil registros.
  - d) Elige un  $k$  razonable, al menos cinco.
  - e) Se puede elegir la definición de distancia libremente.
  - f) Se puede usar cualquier forma razonable para ponderar los votos de los vecinos.
  - g) No se puede usar bibliotecas como `scikit-learn` que implementan el algoritmo.  
Se puede usar bibliotecas como `NumPy` para el cálculo matemático.
3. Sintetiza datos para la clasificación y visualiza los resultados del algoritmo.
4. Explica brevemente el funcionamiento del algoritmo en el informe y muestra el resultado en una figura.
5. Usa la biblioteca de `multiprocessing` para paralelizar el algoritmo.
  - a) No es necesario paralelizar la sintetización de los datos.
  - b) Se tiene que paralelizar el cálculo de distancias y la búsqueda de los vecinos.
  - c) En el informe, explica la estrategia de paralelización.

6. En el caso de clasificar múltiples registros, se puede paralelizar en dos niveles: entre los registros de clasificación y usar un algoritmo paralelo de vecinos cercanos. Explica cual de estos dos tipos de paralelización tendrá un rendimiento mejor.

## Entregables

Entrega un archivo comprimido con el código de Python y un informe corto, elaborado en  $\text{\LaTeX}$ , con las respuestas a las preguntas. Suben el archivo al cuestionario en Siding y usa como nombre del archivo `Tarea1_apellido.zip` en lo cual se tiene que reemplazar `apellido` con tu propio apellido.