



Computación Estadística III
(220257)

Instrucciones:

- Para el presente trabajo, deberán entregar un informe con al menos las siguientes secciones (según corresponda): introducción, índice, índice de figuras, índice de tablas, sección teórica, sección de códigos, conclusión, bibliografía.
- Deberán además adjuntar los archivos .R y .c desarrollados.
- La entrega se hace a través de formulario que se dispondrá en Adecca.
- Fecha de entrega 17 de enero 2023.
- Si el informe se desarrolla en Latex, se obtendrá 5 décimas extras.

Realice lo siguiente:

1. Investigue acerca de las funciones `.C`, `.Call` y `.External` para llamar código C desde R.
2. Suponga que tiene un vector de números y se quiere calcular la suma de los cuadrados de los elementos en el vector.
 - a) Implemente una función R que calcule lo indicado.
 - b) Implemente una función C que calcule lo pedido y pueda ser llamada desde R utilizando el método `.C`.
 - c) Implemente una función C que calcule lo pedido y pueda ser llamada desde R utilizando el método `.Call`.
 - d) Implemente una función C que calcule lo pedido y pueda ser llamada desde R utilizando el método `.External`.
3. El comando `dist` del paquete `stats` de R calcula la matriz de distancia de una matriz de observaciones utilizando la distancia Euclidiana. Sin embargo, esta función puede ser lenta cuando se aplica a un dataset muy grande.
 - a) Investigue acerca del algoritmo que utiliza esta función.
 - b) Busque la implementación que utiliza R de esta función.
 - c) Cree un programa C que implemente de forma mas eficiente esta función, y que pueda ser llamada desde R con alguno de los 3 métodos disponibles.
 - d) Cree un script R que haga uso de lo desarrollado anteriormente en algun dataset de su elección (debe ser lo suficientemente grande) y calcule y compare los tiempos de ejecución.
4. El comando `cor` del paquete `stats` de R calcula la matriz de correlación dado una dataframe o matriz de observaciones. Sin embargo, esta función puede ser lenta cuando se aplica a un dataset muy grande.

- a)* Investigue acerca del algoritmo que utiliza esta función.
- b)* Busque la implementación que utiliza R de esta función.
- c)* Cree un programa C que implemente de forma mas eficiente esta función, y que pueda ser llamada desde R con alguno de los 3 métodos disponibles.
- d)* Cree un script R que haga uso de lo desarrollado anteriormente en algun dataset de su elección (debe ser lo suficientemente grande) y calcule y compare los tiempos de ejecución.