

# Control 4 - IIC2433

Jueves 15 de Octubre de 2020  
Contenidos: Regresiones Logísticas y KNN

---

## Indicaciones

- El control es **individual**. La copia será castigada con nota 1.1 al curso, además de las sanciones disciplinarias correspondientes.
  - El control tiene cuatro preguntas, para obtener el 7 deben contestarlas todas de forma correcta. Cada una tiene su propia ponderación.
  - Para cada pregunta, si esta contiene más de una subpregunta, **debe contestar cada una de las subpreguntas para obtener el puntaje total en la pregunta.**
  - La entrega de este control se realizará a través de un buzón en Canvas, que permanecerá abierto hasta las 23:59 del día Jueves 15 de Octubre. **No se permitirán entregas atrasadas.**
  - El control debe ser entregado en formato **PDF**. En caso de entregar con otro formato, este no se corregirá.
- 

## Preguntas Regresión Logística (2.0 pts)

1. (1.0 pts.) ¿Como se implementa la regresión logística cuando se tienen múltiples clases a predecir? Explique detalladamente el procedimiento para entrenar el modelo y generar las predicciones.
2. (1.0 pts.) ¿Qué propiedades debería cumplir una *squashing function*? Proponga una *squashing function* para reemplazar la sigmoide, y explique cómo separará esta nueva función el espacio de sus datos, puede apoyarse en un dibujo para explicar esto.

## Preguntas KNN (4.0 pts)

1. (2.0 pts.) Un problema de recomendación se define como, dado un usuario y una serie de productos que ha consumido, se busca recomendar algún producto que no ha consumido en base a su historial de compra. Dado este problema, describa como utilizaría el algoritmo KNN para generar recomendaciones de productos a cada usuario, dada una base de datos de transacciones en las que cada fila posee el ID de un usuario, el ID del producto comprado y 2 atributos numéricos que describen el producto comprado (ejemplo: peso y precio). Asuma que todos los usuarios tienen al menos registradas 10 transacciones de productos diferentes en la base de datos. Recuerde que el algoritmo KNN recibe de entrada un conjunto de datos y devuelve una clasificación o regresión, por lo que quizás sea necesario modificar algunos pasos de él.

2. (2.0 pts.) Dado un conjunto de datos de los cuales se sabe que está sujeto a una distribución, la cual es conocida, ¿Qué métrica de distancia será la más adecuada para este problema, si se utiliza el algoritmo KNN? Justifique su respuesta. Ahora, si reducimos las dimensiones simplemente seleccionando columnas para conservar y el resto descartarlas ¿Utilizaría la misma métrica de distancia para KNN? Explique si la mantendría o cambiaría, o si se mantendrá pero con modificaciones, justifique estas decisiones. Recuerde que ahora los datos tendrán menos columnas.