

Control 3 - IIC2433

Jueves 1 de Octubre de 2020
Contenidos: Reglas de asociación

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 1 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 PCA (2.5 pts)

1. (1.5 pts.) Suponga que tiene un conjunto de datos donde ciertas columnas varían sus valores entre 0 y 10.000, y otras entre 0 y 1. ¿Cuáles son las direcciones que tomarán los vectores propios con respecto al conjunto de datos? ¿Por qué? ¿Hay alguna forma de eliminar el sesgo de elección? Si la hay, describa cual.
2. (1.0 pts.) Suponga que tiene dos conjuntos de datos, uno donde quiere entrenar un modelo de predicción, y otro donde quiere calcular predicciones. Si en el proceso de ir mejorando el modelo, se da cuenta que la mejor configuración de él requiere que los datos sean transformados antes con PCA. Indique los pasos a seguir para que el modelo se puede entrenar con el primer conjunto, predecir el segundo, y en qué parte/s se debería aplicar PCA para utilizar la mejor versión del modelo en el proceso de predicción.

Preguntas Regresiones Lineales (3.5 pts)

1. (1.5 pts.) Suponga que tiene un conjunto de datos que tiene solo variables categóricas en los atributos descriptivos. ¿Qué se puede hacer para poder utilizar una regresión lineal sobre este conjunto de datos? Explique al menos dos transformaciones que podría hacer con este conjunto de datos, y cuales serían las ventajas y desventajas de cada una. ¿Cómo afecta cada una de ellas a los parámetros de la regresión lineal?

2. (2.0 pts.) Suponga que tiene dos variables en su conjunto de datos que dependen de una tercera, están correlacionadas. ¿Cómo afecta esto al ajuste de los parámetros de la regresión lineal? ¿Hay algún principio que se esté violando cuando ocurre esto? Por último explique como se relaciona la colinealidad con todo esto.