



Técnicas de Aprendizaje de máquina

Taller 1

Profesor:

Ing. Julio Omar Palacio Niño, M.Sc.

palacio_julio@javeriana.edu.co

Consideraciones

- El taller será desarrollado en máximo 3 estudiantes
- Para el desarrollo del presente taller es necesario disponer de un ambiente de trabajo que incluya un intérprete de Python3.
- La entrega del taller debe ser en un documento PDF con el desarrollo de las preguntas y los archivos que incluyan las implementaciones.

Taller

El objetivo del siguiente taller es hacer un análisis de clasificación

1) Comprensión del dataset

se trabajará con el dataset Statlog (German Credit Data)

2) Comprensión del dataset

- ¿Qué información presenta el dataset?, describir las principales características del dataset, y a su vez identificar la variable sobre la cual se desea hacer la predicción.
- Realice un análisis grafico que permita ver la distribución y ver las correlaciones

3) Limpieza de datos

De las variables que componen el problema cuales pueden ser implementadas para la construcción del modelo, justifique la respuesta.

4) Construcción del dataset

Para realizar el análisis clasificatorio sugiere realizar un particionamiento entre dos conjuntos (entrenamiento, pruebas)

Realice el proceso de construcción del modelo de entrenamiento y pruebas

5) Elaboración del modelo

Para la construcción del modelo se empleará la técnica de clasificador de Bayes, elija el tipo apropiado de clasificador dada la naturaleza del conjunto de datos (tipo de datos de columnas)

Selecciones entre clasificador Gausiano, MultinomialNB, BernoulliNB seleccione uno de ellos

6) Verificación de resultados

- Verifique los resultados del clasificador por medio de matriz de confusión

7) Depuración del modelo

- Emplee un proceso de validación cruzada para evaluar nuevamente el modelo

8) Elaboración segundo modelo

- Seleccione un nuevo modelo entre alguna otra técnica de Bayes o en su defecto un Regresor Logístico, justifique su elección
- Compare los resultados, ¿el nuevo modelo es mejor o peor? / Justifique la respuesta

9) Depuración adicional

- Considere realizar hacer un tratamiento adicional a los datos para ajustar a las características del modelo?
- Justifique su respuesta