

Asignatura Aprendizaje Automático y Análisis de Datos
Guía de Trabajo
Febrero 25 de 2020

Este trabajo debe realizarse de manera individual, y enviarse a la profesora en formato .ipynb a más tardar el martes 3 de marzo antes de iniciar la clase.

Para el conjunto de datos asignado, realizar las siguientes actividades en un archivo jupyter notebook donde se pueda ir documentando y realizando cada paso de manera ordenada:

1. Conocer el conjunto de datos y realizar un plan para ajustar los datos
2. Preprocesar el conjunto de datos (aplicar el plan establecido)
3. Separar el conjunto de datos en conjunto de entrenamiento y de prueba (**)
4. Usar los datos de entrenamiento para construir los siguientes modelos:
 - a. Clasificador por regresión lineal
 - b. Clasificador por regresión logística
 - c. K-vecinos más cercanos ($k=5$)
 - d. Análisis discriminante lineal
 - e. Análisis discriminante cuadrático
5. Calcular la matriz de confusión sobre el conjunto de test para cada uno de los modelos desarrollados
6. Calcular las métricas de desempeño precision, recall y F1-score
7. Analizar los resultados e identificar: fortalezas y debilidades de cada técnica con respecto a este conjunto de datos. Comparar las técnicas y establecer cuál obtiene mejor desempeño

Criterios de evaluación:

Si se comprueba total dominio de la persona sobre el diseño y la implementación presentados se calificará según los porcentajes siguientes, pero si hay desconocimiento de la solución por parte de quien la presenta o

indicios de que no es original, se solicitará el inicio de un proceso disciplinario a la decanatura quien determinará la calificación.

1. Plan de ajuste de los datos: 20%
2. Preprocesamiento: 20%
3. Entrenamiento de los modelos: 15%
4. Resultados obtenidos: 15%
5. Análisis de los resultados: 30%