

Asignatura Aprendizaje Automático y Análisis de Datos
Taller Calificable 1
Marzo 12 de 2021

Objetivo:

Integrar los temas tratados hasta el momento en el curso y aplicarlos a la solución de un problema específico.

Instrucciones:

Este trabajo debe realizarse de manera individual, y enviarse a la profesora mediante el botón de entrega que está después de estas instrucciones en la página de moodle. La fecha límite para entregarlo es el jueves 18 de marzo a las 23:59. Se debe enviar un único archivo en formato .ipynb.

Actividades a realizar:

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 (*Algunos conjuntos de datos ya vienen con los dos archivos separados: .data y .test, en ese caso deben usarse ambos conjuntos y este paso no aplica)
5. Usar los datos de entrenamiento para construir los siguientes modelos:
 1. Clasificador por regresión logística
 2. K-vecinos más cercanos (estimar el mejor valor de k considerando el rango entre $k=3$ y $k=19$)
 3. Análisis discriminante lineal
 4. 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%