



Cursillo Machine Learning
(Jornadas Ingeniería Estadística 2021)

Trabajo grupal a entregar el **29-01-2021**.

Considerar que, al tener una situación en que se clasifica algo como verdadero o falso, tenemos los siguientes conceptos:

Sensibilidad: la proporción de datos verdaderos clasificados como verdaderos sobre el total de datos verdaderos.

Especificidad: la proporción de datos falsos clasificados falsos sobre el total de falsos.

Precisión: la proporción de datos clasificados correctamente sobre el total de datos.

1. Considere el set de datos arrhythmia que puede encontrar dentro de Matlab o adjunto. Contiene dos variables, X con 452 observaciones de 279 características importantes e Y que clasifica las observaciones en 16 clases distintas. El objetivo es desarrollar al menos 3 redes neuronales diferentes (en arquitectura y funciones de activación) que intenten aprender la clasificación entregada. Algunas consideraciones:
 - a) Generar el vector binario que codifica las clases para ser entregado a las redes.
 - b) Para cada red calcular especificidad, sensibilidad y precisión para cada clase, y el promedio total entre todas las clases.
 - c) Utilizar los primeros 400 datos para entrenar. Usar los restantes para medir la generalización de los modelos.
 - d) Indicar cual es el mejor modelo y por que.
2. Proponga un set de datos sobre el que deseen trabajar, estableciendo claramente el objetivo (regresión, clasificación, etc) y desarrolle uno de los modelos vistos en el cursillo para enfrentar el problema.

El entregable es un informe que debe contener al menos lo siguiente:

1. Indices, de contenido, figuras y tablas (si existen).
2. Introducción.
3. Una sección para cada problema, entregando los detalles de los modelos y los resultados.
4. Conclusión.
5. Bibliografía, si corresponde.

Debe entregar además los scripts necesarios para reproducir o visualizar los resultados.

Pueden utilizar cualquier lenguaje de programación que permita el desarrollo de los modelos necesarios.