

INTELIGENCIA ARTIFICIAL

GRADO EN INGENIERÍA DE COMPUTADORES

2023-2024

Práctica 2 (OBLIGATORIA):

1. Objetivo

Esta práctica consiste en aplicar los conceptos vistos relativos a la construcción de modelos de aprendizaje automático, en un contexto médico de forma que el alumno desarrolle su capacidad crítica a la hora de emplear estos modelos.

2. Planteamiento

Se trata de un problema de detección de la benignidad o malignidad de un tumor a partir de imágenes obtenidas por punción de aguja fina. Los datos y su descripción pueden descargarse aquí:

[https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))

3. Tareas a realizar y preguntas

PRIMERA PARTE

1. Dividir la muestra en dos conjuntos (entrenamiento y test) según un criterio razonable (explicar por qué esa división) (0.5 puntos)
2. Realizar alguna comprobación y, en su caso, una intervención razonable sobre los datos (sólo si fuese precisa) para asegurar su completitud, consistencia, etc. (0.5 puntos)
3. Emplear un modelo de k-nn, los hiper-parámetros óptimos mediante algún procedimiento (1 punto).
4. Hacer lo mismo para un modelo lineal empleando algún criterio analítico (1 punto).
5. Hacer lo mismo para un modelo de árboles (1 punto).

6. Hacer lo mismo para un modelo bayesiano empleando como hiper parámetro la constante de suavizado (1 punto).
7. Comparar los resultados obtenidos con los distintos modelos (1 punto)
8. Plantear posibles críticas o mejoras de cada uno de los modelos o del procedimiento seguido, tratando de evitar la repetición de lo mencionado en clase (1 punto)

SEGUNDA PARTE

9. Replicar los análisis anteriores, sobre otro problema médico de características similares (clasificación binaria) explicando **BREVEMENTE** los resultados, empleando un código similar al utilizado en los puntos anteriores. Pueden consultarse bases de datos libres como las contenidas en el repositorio *kaggle.com*. (3 puntos).

Las tareas son interdependientes con excepción de la tarea 9 DE LA SEGUNDA PARTE que puede ser realizada o no por el alumno, al no tener una relación con las anteriores. En caso de no ser realizada la puntuación máxima obtenida será de 7 puntos sobre 10.

4. Formato

La práctica tendrá formato de un Pechakucha, con indicaciones similares a la anterior práctica, en particular deberán entregarse **obligatoriamente** en formato pdf, sin comprimir, las transparencias indicando los nombres y apellidos de los miembros del grupo en la primera página. Adicionalmente es preciso entregar el archivo de código empleado mediante un enlace a *colab*. La carga y manejo de los datos no requerirá de la gestión de ningún archivo (es decir, se empleará **obligatoriamente** la instrucción *from sklearn.datasets import load_breast_cancer*) de forma que el programa permita su ejecución directa. Adicionalmente, para la SEGUNDA PARTE, será preciso proporcionar los datos empleados que, **obligatoriamente**, deberán estar en formato .csv, sin comprimir.

5. Fecha

La práctica se realizará a través de un buzón de entrega en *Blackboard* que estará abierto hasta las **24.00 horas del domingo 12 de noviembre**.