Dpto. Lenguajes y Ciencias de la Computación
E.T.S.I. en Informática, Universidad de Málaga

Aprendizaje Computacional
13 de Diciembre de 2021

Apellidos: **Nombre:**
DNI:

1. Teniendo en cuenta el siguiente conjunto de datos:

x	y	z	Salida
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	1	0

a) Encuentra el nodo raíz de un árbol de Decisión ID3 a partir de la tabla anterior. Teniendo en cuenta que el atributo de decisión es ***Salida*** (2 puntos)
b) Para el siguiente nivel considera un algoritmo de poda tal que se etiqueta el nodo con la clase mayoritaria del atributo objetivo (**Salida**). Completa el árbol. (0.5 punto)
c) A partir de los datos del dataset anterior entrena un árbol de decisión RPART para que prediga el atributo **Salida**. Compara usando validación cruzada (80% para entrenamiento; 20% test) la predicción del árbol obtenido con la de un árbol podado con un CP tal que sea un árbol raíz (1 punto)

2. A partir del conjunto de datos que podemos encontrar en el campus virtual (denominado ejercicio2SVM.csv):
Determina si es separable linealmente e indica cual seria la función Kernel mas adecuada (indica tipo de función y sus parámetros) (0,5 puntos).
Calcula los siguientes parámetros de la Maquina de Soporte Vectorial que podemos obtener con el dataset anteriores y el Kernel elegido:

I. Vectores Soporte. (0.25 puntos)
II. Ancho del canal (0.5 puntos)
III. Vector de Pesos normal al Hiperplano (W) (0.25 puntos)
IV. Vector B (0.25 puntos)
V. La ecuación del Hiperplano y de los planos de soporte positivo y negativo. (0.75 puntos)
VI. Pinta el conjuntos de puntos y el Hiperplano. (0.75 puntos)
VII. Clasifica los puntos (0.5, 0,8) y (0.6,0,2). (0.25 puntos)

3. La URL maliciosa, también conocido como sitio web malicioso, es una amenaza común y grave para la ciberseguridad. La URL maliciosa atraen a los usuarios confiados a ser victimas de estafas (perdidas monetarias, robo de información privada, e instalación de malware), y causan perdidas de miles de millones de euros cada año. Por tanto, es necesario detectar y actuar sobre tales amenazas de manera oportuna. Tradicionalmente, esta detección se realiza principalmente mediante el uso de listas negras.. Sin embargo, las lista negras no pueden ser exhaustivas, y carecen de la capacidad para detectar nuevas

Dpto. Lenguajes y Ciencias de la Computación
E.T.S.I. en Informática, Universidad de Málaga

Aprendizaje Computacional
13 de Diciembre de 2021

Apellidos: **Nombre:**
DNI:

URL maliciosas generadas. Otra técnica que se esta usando es Machine Learning para clasificar las URL entre maliciosas o no. Se pide:
a) Elige el clasificador que mejor predicción produzca (usa el accuracy y validación cruzada para entrenamiento y predicción) entre SVM y Perceptron Multicapa. El dataset *detect-maliciusus-URL.csv* se encuentra en el campus virtual. (3 puntos).