

MINERIA DE DATOS UTILIZANDO SISTEMAS INTELIGENTES

PRACTICA 5 – PERCEPTRÓN Y MULTIPERCEPTRON

Material de Lectura: Capítulo 13 del Libro Introducción a la Minería de Datos de Hernández Orallo

Ejercicio 1

El archivo **Iris.xls** contiene información referida a la longitud y al ancho de sépalos y pétalos de tres especies de flores, *iris setosa*, *iris versicolor* e *iris virginica*.

- Utilice los datos del archivo **Iris.xls** para entrenar un perceptrón que permita determinar si una flor pertenece a la especie *iris setosa* o no. Utilice un operador “*Generate Attributes*” para generar un nuevo atributo llamado **clase** que tenga el valor “Iris_setosa” si el dato pertenece a la clase *iris setosa* o el valor “Otra” en caso contrario. Luego utilice un operador “*Set Role*” para indicar que este nuevo atributo es el atributo que contiene la clase. Verifique que el modelo obtenido clasifica correctamente todos los patrones de entrenamiento utilizando los operadores “*ApplyModel*” y “*Performance*”.
- El archivo **IrisTest.xls** contiene información de cuatro flores sin clasificar. Utilizando Rapid Miner, indique para cada una de ellas la respuesta del perceptrón completando la columna **clase** a fin de determinar en cada caso si la flor pertenece a la especie *iris setosa* o no.
- Rehacer manualmente el inciso b) utilizando los valores del modelo correspondiente al perceptrón entrenado en a). Verifique que la clasificación obtenida en este caso coincide con la del inciso b). Incluya en su respuesta todos los cálculos realizados.

Ejercicio 2

El archivo **Vinos.xls** tiene información referida a 13 características químicas y/o visuales de varias muestras de vinos pertenecientes a 3 clases distintas.

- Mediante el operador Split Data, utilice el 80% de los datos del archivo **Vinos.xls** para entrenar un multiperceptrón (operador “*Neural Net*”) que sea capaz que distinguir entre las 3 clases de vinos. Observe la tasa de acierto obtenida sobre el 20% restante.
- Mediante el operador Split Data, utilice el 80% de los datos del archivo **Vinos.xls** para entrenar un árbol de decisión C4.5 (operador “*W-J48*”) que sea capaz que distinguir entre las 3 clases de vinos. Observe la tasa de acierto obtenida sobre el 20% restante.
- Compare los resultados obtenidos en a) y en b) con respecto a la tasa de acierto obtenida y en cuanto a la facilidad de interpretar el modelo obtenido. ¿En qué circunstancias utilizaría cada uno de ellos?

Ejercicio 3

El archivo **Balance.xls** tiene información sobre un experimento psicológico realizado para evaluar el aprendizaje en los niños. Cada fila de la tabla tiene las características de una balanza, referidas a la longitud de los brazos izquierdo y derecho de la balanza y al peso que hay en cada brazo, y un atributo que indica si la balanza se inclina al lado izquierdo (L), derecho (R), o está balanceada (B).

Rehaga lo solicitado en el Ejercicio 2 pero utilizando esta vez los datos del archivo **Balance.xls**.

Ejercicio 4

Se ha entrenado un Perceptrón para que sea capaz de distinguir hojas de HELECHO y de TREBOL a partir de dos características medidas sobre su imagen: área y perímetro. El modelo obtenido fue:

INTERCEPT= 1.181647, $W(\text{Perímetro}) = 1011.642$ y $W(\text{Area}) = -134.399$.

Si un ejemplo cuyos valores son Perímetro = 700 y Area = 5000 corresponde a una hoja de HELECHO ¿cómo clasifica el Perceptrón entrenado a un ejemplo cuyos valores son Perímetro = 650 y Area = 5500?

Ejercicio 5

El archivo **Frutas.xls** contiene información referida al diámetro y el color de naranjas y melones.

a) Utilice todos los ejemplos de la hoja **FrutasNumerico** para entrenar un perceptrón que separe los ejemplos en estas dos clases. Luego realice un diagrama de dispersión y grafique el modelo (recta) obtenido al finalizar el entrenamiento.

b) Utilice todos los ejemplos de la hoja **FrutasNumerico** para construir un árbol mediante el método C4.5 (operador W-J48) que separe los ejemplos en estas dos clases. Luego realice un diagrama de dispersión y grafique los distintos cortes realizados por el modelo (árbol) en el rango de cada atributo para realizar la clasificación.

c) Utilice todos los ejemplos de la hoja **FrutasNominal** para construir reglas que permitan asociar los tres atributos (Diámetro, Color y Clase). Elija las dos mejores reglas de asociación que obtuvo para cada clase.

d) Si tuviera que elegir uno de los tres modelos, ¿cuál elegiría y por qué?