

Aprendizaje automático



Tutorial 3

Grupo 84

Sabrina Riesgo Reyes
Laura Yunta García

100363834
100363785

100363834@alumnos.uc3m.es
100363785@alumnos.uc3m.es

Índice

1. INTRODUCCIÓN.	3
2. EJERCICIOS DEL TUTORIAL.	3
a. Ejercicio 1.	3
i. Pregunta 11.	3
ii. Pregunta 13.	3
b. Ejercicio 2.	3
i. Pregunta 3.5.	4
ii. Pregunta 3.6.	4
iii. Pregunta 3.7.	4
iv. Pregunta 3.9.	4
3. CONCLUSIONES Y DIFICULTADES.	4

1. INTRODUCCIÓN.

El desarrollo de la práctica busca adquirir los conocimientos y familiarizarse con dos interfaces de Weka que facilitan el diseño y la ejecución de múltiples experimentos de una sola vez ya que las interfaces Experimenter y KnowledgeFlow permiten ejecutar distintos algoritmos de aprendizaje automático sobre distintos conjuntos de datos y comparar los resultados usando test estadísticos. Finalmente, el documento y con ello la práctica se dividirá en tres partes diferenciadas:

- **Introducción:** En este apartado se comenta brevemente el objetivo a alcanzar con el desarrollo del tutorial y se explica la estructura y contenido de cada uno de los apartados del documento.
- **Respuestas a los ejercicios:** En este apartado se responderán a cada una de las preguntas especificadas en el documento del tutorial.
- **Conclusiones y problemas encontrados:** En este apartado se comentarán las conclusiones sacadas con la realización del ejercicio, además los diversos problemas que se han conseguido solventar durante el transcurso de la práctica.

2. EJERCICIOS DEL TUTORIAL.

En este subapartado se responderán a todas las preguntas mencionadas en el enunciado de la práctica. Cabe destacar que los diversos puntos de los ejercicios a responder que no tienen preguntas (es decir que solo especifican los pasos a realizar para la realización) no se contemplarán en el documento y únicamente se podrán visualizar aquellos puntos que tienen pregunta.

a. Ejercicio 1.

i. Pregunta 11.

Ejecutar el flujo de conocimiento. Para ellos ejecutar la opción Start loading de nodo Arff loader. Seleccionar la opción Show Results en los nodos Text Viewer. ¿Qué se muestra en cada uno de ellos? ¿Cuál es el porcentaje de instancias clasificadas correctamente?

En el caso del nodo TextViewer que se encuentra directamente conectado al nodo J48, se muestra el árbol y finalmente, el número de hojas y el tamaño del árbol. Sin embargo, en el caso del nodo TextViewer que se encuentra conectado al nodo J48 mediante un nodo Classifier PerformanceEvaluator, se muestra el porcentaje de instancias bien y mal clasificadas entre otros porcentajes, el total de instancias, detalles sobre la precisión y la matriz de confusión. El porcentaje de instancias clasificadas correctamente es 86.2105%.

ii. Pregunta 13.

¿Cuál es la utilidad de crear flujos de conocimientos con esta interfaz de Weka?

Ofrece soporte para el aprendizaje incremental, describe mejor el funcionamiento interno del programa, permite realizar pruebas de una forma sencilla y comparar algoritmos y conjuntos de datos.

b. Ejercicio 2.

Para la realización del siguiente ejercicio se han utilizado tres ficheros distintos de atributos. El primero se trataría del fichero base que contiene todos los atributos, a continuación se puede

encontrar el fichero que tiene los atributos relacionados con los fantasmas del juego y por último, se puede visualizar el fichero que contiene los atributos relacionados con el Pac-Man.

i.Pregunta 3.5.

¿Hay algún conjunto de datos que parezca más adecuado?

Tras haber realizado el experimento comparando los tres ficheros de atributos se puede concluir que el más adecuado es el fichero del primer subconjunto creado con los atributos de los fantasmas. Debido a que es el conjunto de datos que es mejor en más ocasiones dependiendo del algoritmo a utilizar.

ii.Pregunta 3.6.

¿Qué algoritmo te parece más adecuado?

Se considera el algoritmo más adecuado tras realizar el experimento al algoritmo IBk -1 ya que garantiza en más ocasiones que el conjunto de datos sea mejor, siendo el único que garantiza que sean dos conjuntos los mejores.

iii.Pregunta 3.7.

¿Son los resultados del mejor algoritmo mucho mejor que los del resto?

Consideramos que no se puede afirmar que el algoritmo mejor es mucho mejor que el resto en todas las ocasiones. Sí que es cierto que comparándolo con algoritmos como Naive Bayes es mucho mejor el algoritmo IBk-1 pero si lo comparamos con el algoritmo IBk-3 no se aprecia una gran diferencia.

iv.Pregunta 3.9.

¿Por qué o para qué te parece adecuado el uso de Experimenter de Weka?

El uso del Experimenter de Weka permite comparar sistemáticamente los algoritmos predictivos de Weka sobre diferentes conjuntos de datos.

3. CONCLUSIONES Y DIFICULTADES.

Una vez realizada la práctica podemos concluir que es de gran importancia la buena elección de los algoritmos predictivos ya que dependiendo de cuál se use en cada caso pueden salir unos resultados mejores o peores.

Cabe destacar que los problemas encontrados en la realización de la práctica no han tenido gran trascendencia debido a que se trataba de una dificultad inicial a la hora de interpretar los resultados obtenidos con la ejecución de Weka, que finalmente se han solventado a medida que la íbamos realizando y adquiriendo conocimiento sobre el funcionamiento de la misma ya que en un principio el tutorial no especifica demasiado la interpretación de los resultados.

Finalmente, hemos tenido problemas a la hora de generar las cabeceras con el fichero de salida de Pac – Man y con el formato de fichero que se iba introducir en Weka pero finalmente conseguimos solucionarlo.