



Práctica

Presentación

Práctica sobre el desarrollo y aplicación de algoritmos de aprendizaje.

Competencias

Competencias de grado

- Capacidad de utilizar los fundamentos matemáticos, estadísticos y físicos y comprender los sistemas TIC.
- Capacidad para analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para abordarlo y resolverlo.
- Capacidad para conocer las tecnologías de comunicaciones actuales y emergentes y saberlas aplicar convenientemente, para diseñar y desarrollar soluciones basadas en sistemas y tecnologías de la información.
- Capacidad para proponer y evaluar diferentes alternativas tecnológicas y resolver un problema concreto.

Competencias específicas

- Capacidad para utilizar la tecnología de aprendizaje automático más adecuada para resolver un determinado problema.
- Capacidad para evaluar el rendimiento de los diferentes algoritmos de resolución de problemas mediante técnicas de validación cruzada.

Objetivos

Siguiendo con los datos utilizados en las PECs, disponemos de un conjunto de datos de clientes de una distribuidora de alimentos (archivos “Wholesale customers.csv” y “Wholesale customers.txt”). Cogeremos como clase la primera columna (*channel*), que representa si el cliente compra al detalle (*retail*) o al por mayor (*horeca*). Se quiere desarrollar un clasificador a partir de las diferentes variables (6 en total, ver la descripción en el archivo “Wholesale customers.txt”) para ver si



podemos predecir el tipo de cliente. Notad que tenemos que ignorar la segunda columna.

Queremos construir un modelo capaz de predecir la clase del ejemplo utilizando el algoritmo de clasificación probabilístico Naïve Bayes. El objetivo es implementar el algoritmo y evaluar los resultados con los datos comentados.

La práctica está formada por 3 ejercicios con diferente peso. El informe de la práctica tiene que contener las respuestas de todos los ejercicios.

Descripción de la Práctica

Ejercicio 1. Naïve Bayes (20%)

El Naïve Bayes es un algoritmo de clasificación probabilístico pensado para trabajar con conjuntos de datos con atributos nominales. Para poder trabajar con datos numéricos tenemos que hacerle una pequeña modificación.

Buscad información sobre el “Gaussian Naïve Bayes” y describidlo en detalle, poniendo énfasis en cómo se tiene que implementar.

Ejercicio 2. Weka (20%)

Utilizando el Weka aplicad **dos** algoritmos diferentes de aprendizaje supervisado (uno de ellos el Naïve Bayes) para predecir la clase del archivo de test. Dad la precisión (*accuracy*) de la clasificación obtenida con cada método. Discutid los resultados.

Ejercicio 3. Implementación del Naïve Bayes (60%)

Implementad el algoritmo del Naïve Bayes para predecir si los ejemplos del conjunto de test son *retail* u *horeca*, tratando directamente los atributos numéricos.

Tened en cuenta que el informe tiene que contener:

1. Una explicación del algoritmo implementado, explicitando los detalles que consideréis relevantes y las decisiones de diseño que hayáis tomado. Mencionad especialmente los pasos de la implementación del algoritmo.



2. Una tabla con al menos la precisión, la matriz de confusión y el tiempo de cálculo del algoritmo, comparando los resultados con los obtenidos con Weka (ejercicio 2).
3. Un apéndice con el listado del código fuente del programa.
4. Y, en general, una justificación de todo lo que estéis haciendo.

Tenéis que entregar el programa que hayáis implementado para realizar el ejercicio 3 (se recomienda que sea en Python o Java, pero la podéis realizar en Delphi, Lazarus, C, C++, Visual Basic o “similar”). La calidad del código (estructura, comentarios...) es uno de los criterios importantes de corrección.

Recursos

Básicos

Para realizar esta práctica disponéis de unos archivos adjuntos (“Wholesale customers.txt”, “Wholesale customers.csv”) donde encontraréis los datos correspondientes.

Criterios de valoración

Los tres ejercicios de esta práctica se valoraran con 2, 2 y 6 puntos respectivamente, repartidos de la forma siguiente:

Ejercicio 1: este ejercicio vale 2 puntos que valorará significativamente la claridad, simplicidad y estructura del discurso; así como el enfoque de cara la implementación del mismo.

Ejercicio 2: se valoraran la inclusión de la tabla de resultados con 1,5 puntos. Los resultados tienen que contener como mínimo la precisión (*accuracy*), el número de ejemplos erróneos y las matrices de confusión de todos los métodos especificados en el enunciado. Los 0,5 puntos restantes se adjudican a los comentarios, valoraciones y justificaciones de todo el ejercicio.

Ejercicio 3: la valoración de este ejercicio va por fases:

- El funcionamiento del programa tienen una valoración binaria de 0 o 4 puntos en función de si funciona o no.
- La calidad del código y la presentación se llevan un punto. Como calidad entendemos: la estructura, comentarios,



diseño...; y como presentación: la salida del programa, el uso de entornos de ventanas, la entrada de los datos, la posible parametrización del programa...

- El punto restante valorará: las conclusiones generales, el análisis de resultados, las comparaciones con el Weka...

Formato y fecha de entrega

Tenéis que entregar la práctica en un único archivo comprimido que contenga la memoria en formato *pdf* y el código fuente del programa.

El nombre del archivo tiene que ser ApellidosNombre_AC_Practica con extensión .zip.

Fecha límite: 29 de mayo a las 24 horas.

Para dudas o aclaraciones sobre el enunciado, dirigíos al consultor responsable de vuestra aula.

Nota: Propiedad intelectual

A menudo es inevitable, al producir una obra multimedia, hacer uso de recursos creados por terceras personas. Es por tanto comprensible hacerlo en el marco de una práctica de los estudios del Grado de Informática, siempre que esto se documente claramente y no suponga plagio en la práctica.

Por lo tanto, al presentar una práctica que haga uso de recursos ajenos, se presentará junto con ella un documento en el que se detallen todos ellos, especificando el nombre de cada recurso, su autor, el lugar donde se obtuvo y el su estatus legal: si la obra está protegida por copyright o se acoge a alguna otra licencia de uso (Creative Commons, licencia GNU, GPL ...). El estudiante deberá asegurarse de que la licencia que sea no impide específicamente su uso en el marco de la práctica. En caso de no encontrar la información correspondiente deberá asumir que la obra está protegida por copyright.

Deberán, además, adjuntar los archivos originales cuando las obras utilizadas sean digitales, y su código fuente si corresponde.