



Primera Práctica Aprendizaje Automático

Análisis de Mercado de Ordenadores Portátiles

4º Ingeniería en Informática
Curso 2005-06 (Junio)

1. Introducción

La oferta de ordenadores portátiles proporciona un conjunto de datos que puede ser interesante tratar con el objetivo de construir un sistema que de forma automática permita clasificar dichos ordenadores según sus características, o estimar el precio más adecuado para un determinado ordenador. Este tipo de sistema puede resultar interesante tanto desde el punto de vista del comprador del portátil, como desde el punto de vista del vendedor:

- Estimar el precio de un producto es una tarea que debe realizar tanto la persona que pretende realizar una venta de la misma, como el comprador. Para establecer el coste de un producto es necesario tener en cuenta tanto sus características como el precio de mercado que tienen los productos con unas características similares.
- Clasificar un producto en distintas clases que reflejen la relación calidad-precio del producto. Por ejemplo, si tiene un precio normal, si es demasiado caro, o por el contrario es un chollo, es más útil desde el punto de vista del comprador.

En esta práctica se pretende analizar la utilidad de las técnicas de aprendizaje automático estudiadas en teoría, aplicándolas a un dominio como el de la clasificación de ordenadores portátiles, bien sea con el objetivo de estimar su precio, o bien con el objetivo de clasificar el producto en clases relativas a lo adecuado de su precio. Trabajaremos con los dos casos.

2. Descripción de la Tarea

Una parte muy importante y en muchos casos laboriosa que hay que llevar a cabo para construir cualquier sistema utilizando aprendizaje automático es conseguir un conjunto de datos apropiado para entrenarlo. Para conseguir estos datos de entrenamiento es necesaria una fuente de información que los proporcione. Además, una vez que se cuenta con los datos habrá que analizarlos y, posiblemente, preprocesarlos y hacer pruebas hasta conseguir un conjunto de datos *adecuado*.

Para llevar a cabo la práctica el alumno deberá elegir varios sitios Web o fuentes de información similares en el que se ofrezcan datos sobre ordenadores portátiles a la venta. Esta será la fuente de información, a partir de la cual deberá obtener los datos tanto para entrenar el clasificador, como para evaluarlo.

Entre las características que pueden ayudar a definir el precio de un ordenador portátil pueden encontrarse las siguientes:

- El procesador.
- La memoria RAM.
- El tamaño y velocidad del disco duro

- Tipo de pantalla
- Peso
- Etc.

Aunque aquí se han definido algunas características posibles, queda a decisión del alumno la selección del conjunto definitivo, pudiendo eliminar o incluir otros atributos de los definidos anteriormente.

El correcto desarrollo de la práctica necesitará los siguientes pasos:

1. Capturar los ejemplos de entrenamiento en formato WEKA, a partir del contenido de páginas Web y/o distintas fuentes de información.
 - Creación de un directorio llamado *CasosEntrenamiento* que almacenará todos los ejemplos que se utilicen para entrenar al clasificador, es decir, las páginas Web que describan los ordenadores.
 - Decidir qué atributos de los ordenadores debe contener cada ejemplo de entrenamiento.
 - A partir del contenido del directorio *CasosEntrenamiento* y de los atributos que se decidan en el apartado anterior, generar el conjunto de datos de entrenamiento en formato WEKA. Este paso puede realizarse de forma manual o de forma automática. Si se implementa un método automático, la única restricción que tiene que cumplir el software que se desarrolle es que pueda ser ejecutado en las aulas del laboratorio. El tamaño del conjunto de entrenamiento debe ser decidido por el alumno
2. Utilizar WEKA para construir los sistemas para las tareas que se han comentado anteriormente. En el caso de estimar el precio de un portátil, la clase será numérica, o discretizada en un rango de valores *adecuado*. En el caso de la clasificación de ordenadores según su relación calidad-precio, la clase será discreta.
 - Seleccionar cuáles son los atributos más relevantes.
 - Utilizar para cada tarea (cálculo de precio y clasificación) cualquier clasificador disponible en la herramienta WEKA.
3. Repetir los pasos anteriores tantas veces como sea necesario, hasta encontrar el mejor clasificador, con el conjunto de atributos adecuado.

3. Entrega de la Práctica

La práctica se debe entregar el día **21 de abril de 2005**.

Para realizar la entrega deberéis:

- Entregar un CD en el que aparezcan los nombres de los integrantes del grupo y en el que se incluyan los siguientes ficheros:
 - Directorio de páginas Web de entrenamiento comprimido `entrenamiento.tgz`
 - Los ficheros con los datos de entrenamiento en formato WEKA. En caso de que se implemente un método automático de creación de ficheros en formato WEKA, habrá que incluir tanto los ficheros fuentes como el ejecutable y los ficheros necesarios para generar el ejecutable (`makefile`) si procede.
 - El fichero de entrenamiento definitivo, tras la aplicación de filtros y/o selección de características.
 - El Modelo o clasificador generado.
- Entregar una memoria (de unas 15-30 páginas) en papel con la siguiente estructura:
 1. **Introducción** (1 hoja de descripción de la práctica desde vuestro punto de vista)

2. **Análisis de atributos y Generación de Datos:** Análisis inicial de los atributos seleccionados para caracterizar los ordenadores. En caso de que se implemente, descripción del método automático para la generación de los datos de entrenamiento en formato WEKA.
3. **Construcción del clasificador:** Descripción de las pruebas que se han realizado hasta conseguir el clasificador final y descripción del clasificador conseguido. El hecho de que se realicen pruebas puede suponer cambios en los atributos seleccionados, tipo de representación o cualquier otro. Estos cambios, junto con las decisiones tomadas, deben ser explicados en este apartado.
4. **Análisis de resultados:** Análisis de los resultados sobre el conjunto de entrenamiento. Análisis acerca de la calidad real del clasificador.
5. **Conclusiones:** En las que además de cualquier tipo de conclusión a la que se llegue tras la realización de la práctica, se debe incluir:
 - Opinión acerca del proceso de construcción de un clasificador. ¿Cuáles son las principales dificultades que has encontrado?.
 - Opinión acerca de la aplicación de este tipo de aprendizaje automático a problemas reales. ¿Se te ocurren otras aplicaciones?.
 - Opinión sobre la conveniencia de la práctica, de lo que se ha aprendido, ... etc.

Las conclusiones son una parte importante de la práctica, no las descuidéis.