

# Selección de Atributos en Métodos no Lineales de Aprendizaje de Preferencias para el Análisis de Datos Sensoriales

Gustavo Fernández Bayón

Centro de Inteligencia Artificial  
Universidad de Oviedo

**Resumen** La calidad de los productos alimenticios puede ser valorada de diferentes formas. En este texto nos centramos en aquellos aspectos obtenidos a partir de impresiones sensoriales. Cuando se trata de inducir una función capaz de evaluar descripciones de los objetos, se debe tener en cuenta que las valoraciones de los consumidores no son más que una forma de expresar sus preferencias relativas acerca de los objetos juzgados en una única sesión. El aprendizaje de preferencias es propuesto como alternativa a los métodos clásicos de regresión. Se presentan a su vez dos métodos de selección de atributos especialmente adaptados. Las pruebas de estas aproximaciones se llevan a cabo sobre dos bases de datos sensoriales reales.

## 1. Introducción

Una parte importante del éxito de la industria alimentaria se basa en su habilidad para crear productos que satisfagan las necesidades sensoriales de los consumidores. De esta forma, es necesario organizar encuestas para descubrir las opiniones de los potenciales consumidores sobre la calidad o aceptación de los productos a comercializar, cualidades éstas que son percibidas principalmente a través de sensaciones. El objetivo del análisis de datos sensoriales es procesar esas impresiones de los consumidores, las cuales pueden ser representadas en el marco de los problemas de regresión: la descripción de de cada objeto  $x$  en un conjunto  $E$  es acompañada de un número  $r(x)$  que indica el grado de satisfacción de cada consumidor, o el valor medio para un grupo de consumidores.

Tradicionalmente, el proceso seguido para trabajar con estos conjuntos de ejemplos incluye la prueba de hipótesis estadísticas [1,2,3], así como aproximaciones basadas en redes bayesianas [4]. Estos enfoques requieren que *todos* los productos alimenticios (los objetos  $x$ ) deben ser evaluados por *todos* los consumidores; en la práctica, esto es imposible la mayoría de las veces. En general tendremos conjuntos de valoraciones  $(r(x) : x \in E_i)$  para cada consumidor o grupo de consumidores  $i$ , donde  $\cup(E_i : i \in I) = E$ .

Otra alternativa puede ser el enfoque del problema mediante técnicas de regresión. De esta manera, se puede intentar inducir una función que, tomando

como entrada la descripción de un objeto, nos devuelva su valoración. Sin embargo, esta manera de modelizar el conocimiento de los expertos no es fiable. De hecho, los algoritmos de regresión frecuentemente obtienen errores cercanos a los conseguidos por el *predictor de la media* trivial, esto es, el predictor que devuelve la media de  $\cup((r(x) : x \in E_i) : i \in I)$  como la valoración de todos los objetos. Esto claramente no es realista cuando se trata de descubrir las preferencias de los usuarios.

En la siguiente sección se presentarán las peculiaridades del trabajo con datos sensoriales. De forma especial se quiere hacer hincapié en el hecho de que las preferencias de los consumidores representan de forma natural una ordenación relativa de los objetos presentados. Existe un *efecto lote*<sup>1</sup>, que a menudo produce un sesgo en las valoraciones. Esto implica que se deberá considerar, como un aspecto muy importante, la información acerca de los lotes presentados a los usuarios durante el transcurso de una sesión de valoración.

Posteriormente se presentará otra forma de enfrentarse al análisis de datos sensoriales en Aprendizaje Automático, basada en el aprendizaje de preferencias [5,6,7]. En este contexto, los ejemplos de entrenamiento pueden ser representados como juicios de preferencias. Como se explica en [8,9], la obtención de información en forma de juicios de preferencias puede ser más fácil y natural que la obtención de las valoraciones necesarias para una aproximación por regresión.

En la sección siguiente se describirán dos métodos de selección de atributos especialmente adaptados para sistemas de aprendizaje no lineales, así como un filtro diseñado para combatir la redundancia presente en este tipo de problemas.

Finalmente se mostrarán los resultados de estas aproximaciones en los experimentos realizados sobre dos bases de datos de carácter sensorial, obtenidas a partir de los grados de aceptación de la carne de vacuno y la sidra tradicional asturiana respectivamente.

## 2. Análisis de Datos Sensoriales

Un muy buen informe acerca del uso de los datos sensoriales en la industria alimentaria puede encontrarse en [2,3]. En [4] y en [10] se aborda el mismo tema desde la perspectiva del Aprendizaje Automático.

Desde un punto de vista conceptual, necesario para una aproximación basada en técnicas de Aprendizaje Automático, los datos sensoriales incluyen la valoración de productos alimenticios determinada por dos tipos de grupos de personas, normalmente llamados *paneles*. El primer grupo está formado por un número pequeño de expertos entrenados; éstos describen los productos mediante pares atributo-valor. A estos expertos se les requiere tener la suficiente precisión como para ser capaces de discriminar productos muy similares. Nótese que a los expertos no se les tiene por qué preguntar explícitamente acerca de la calidad o aceptación global del producto, ya que su opinión puede ser bastante diferente a la de un usuario no entrenado.

---

<sup>1</sup> Distorsión sensorial por la que un objeto rodeado por otros peores obtiene mejor puntuación de un experto que si estuviera rodeado de objetos mejores.

El segundo tipo de panel está formado por un grupo de consumidores desentrenados. Estos tienen que valorar su grado de aceptación o satisfacción respecto a los productos presentados. El objetivo es ser capaces de relacionar descripciones sensoriales (humanas y mecánicas) con las preferencias del consumidor para mejorar las decisiones de producción.

Los productos alimenticios presentan una importante peculiaridad: los consumidores no pueden probar todas las muestras disponibles. En algunos casos es imposible físicamente, y en otros una prueba tan grande podría mermar la capacidad sensorial del consumidor. Normalmente, un consumidor sólo participa en un pequeño número de sesiones de prueba, y éstas suelen producirse el mismo día.

Una demostración de la importancia de la información acerca de las sesiones puede verse de forma gráfica en la Figura 1. En ella se puede ver un conjunto de valoraciones de consumidores (representadas en el eje vertical) acerca de varios productos cuyas descripciones vienen dadas por un único valor real  $x$ , representado en el eje horizontal. Si se observa la imagen de la izquierda, donde las valoraciones correspondientes a la misma sesión están rodeadas por una elipse, se puede concluir que, en todas las sesiones, el mensaje del consumidor es el mismo: a mayor valor de  $x$ , mayor valoración del objeto. Sin embargo, existen discrepancias en la expresión de este conocimiento a lo largo de las sesiones, seguramente debido a que cada sesión está formada por diferentes consumidores; o quizás porque el mismo consumidor olvida el valor exacto que determina un nivel determinado de satisfacción; o porque las reacciones sensoriales varían de una sesión a otra.

Si no se considera la información relativa a las sesiones, los datos recogidos se convierten en la nube de puntos de la gráfica derecha de la Figura 1. De esta forma es difícil para un método de regresión descubrir la opinión -unánime- de los consumidores. De hecho, en este caso, los métodos de regresión concluyen que, a mayor  $x$ , peor valoración del producto, ya que esa parece ser la orientación general de la nube de puntos en el espacio. De todos modos, nótese que cualquier predicción hubiera sido posible con tan sólo algún cambio en las posiciones relativas entre sesiones.

### 3. Aprendizaje de Preferencias

El objetivo es construir un conjunto de juicios de preferencias de la forma

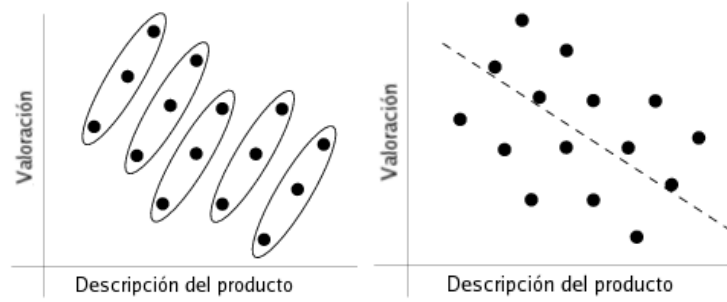
$$PJ = \{v_j > u_j : j = 1, \dots, m\} \quad (1)$$

Siguiendo la aproximación descrita en [11,5], se busca una función  $F : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  tal que

$$\forall x, y \in \mathbb{R}^m, F(x, y) > 0 \Leftrightarrow F(x, 0) > F(y, 0) \quad (2)$$

De esta manera la función de valoración  $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$  puede ser definida como

$$\forall x \in \mathbb{R}^m, f(x) = F(x, 0) \quad (3)$$



**Figura 1.** La gráfica de la izquierda representa, dentro de cada elipse, las valoraciones de las diferentes sesiones. Si no se tiene en cuenta la información acerca de las sesiones, la nube de puntos de la imagen de la derecha representa el conjunto de datos de entrada para los métodos de regresión. Esta interpretación de los datos aconseja buscar otro tipo de enfoques del problema.

Dado el conjunto de juicios de preferencias PJ (Ecuación 1), se puede caracterizar  $F$  por medio de las siguientes restricciones

$$\forall j = 1, \dots, m, F(v_j, u_j) > 0 \text{ y } F(u_j, v_j) < 0 \quad (4)$$

Queda definido así otro problema de clasificación binaria que puede ser resuelto por una SVM, dando como resultado una función de la forma

$$F(x, y) = \sum_{i=1}^n \alpha_i z_i \mathbb{K}(x_i^{(1)}, x_i^{(2)}, x, y) \quad (5)$$

donde los pares  $(x_i^{(1)}, x_i^{(2)})$  son los vectores de soporte, y  $\mathbb{K}$  es el kernel usado por SVM. La idea principal es la definición de este kernel como

$$\mathbb{K}(x_1, x_2, x_3, x_4) = k_p(x_1, x_3) - k_p(x_1, x_4) - k_p(x_2, x_3) + k_p(x_2, x_4) \quad (6)$$

Aquí,  $k_p$  es una función kernel definida como el producto escalar de las representaciones de dos objetos. Nótese que, de acuerdo a las anteriores definiciones, tenemos

$$f(x) = \sum_{i=1}^n \alpha_i z_i (k_p(x_i^{(1)}, x) - k_p(x_i^{(2)}, x)) \quad (7)$$

Como se puede observar, cuando se usa el kernel polinómico se obtiene una función de valoración no lineal que asigna un valor a cada objeto  $x$ .

#### 4. Selección de Atributos

Un aspecto a tener en cuenta cuando se trabaja con problemas reales de análisis sensorial es la relevancia de los atributos. Se hace necesario encontrar

aquellos atributos que tienen más influencia en los gustos de los consumidores. Así, los procesos productivos se pueden centrar en ellos para mejorar la aceptación del producto final. Lo que es más, la reducción del número de variables que describen a un objeto disminuye el coste de adquisición de los datos, lo que en muchos casos hace que estas técnicas sean aplicables en el ámbito industrial [10].

Durante los últimos años, varios métodos de selección de atributos para SVM han sido desarrollados. Uno de los más importantes es RFE [12]. Este método genera una ordenación de los atributos que describen los objetos en función de su utilidad de cara a aprender el problema.

No obstante, el criterio de RFE no es aplicable directamente a kernels no lineales. Debido a esto se han usado dos métodos reconocidos para la obtención de ordenaciones de atributos en escenarios no lineales. Además, se debe tener en cuenta una peculiaridad de los conjuntos de juicios de preferencias. En este caso, los ejemplos están formados por pares de objetos  $(v, u)$ , donde los dos objetos son descritos por el mismo conjunto de  $d$  atributos. Por lo tanto, se debe considerar cada atributo como doble a la hora de evaluar su utilidad; y deshacerse de las dos copias en caso de eliminación.

#### 4.1. Métodos de ordenación de atributos

*Método 1.*- El primer método aplicado para obtener una ordenación de atributos con kernels no lineales fue propuesto por A. Rakotomamonjy en [13]. Su criterio ordena la lista de atributos en función de su influencia en los pesos de los atributos. De hecho, es una extensión de RFE al caso no lineal. En cada iteración elimina el atributo que minimiza el siguiente valor

$$R_1(i) = |\nabla_i \|\mathbf{w}\|^2| = \left| \sum_{k,j} \alpha_k \alpha_j z_k z_j \frac{\partial K(\mathbf{s} \cdot \mathbf{x}_k, \mathbf{s} \cdot \mathbf{x}_j)}{\partial s_i} \right|, \quad i = 1, \dots, d \quad (8)$$

donde  $s$  es un factor de escala usado para simplificar el cálculo de las derivadas parciales.

*Método 2.*- El segundo método ha sido usado en [14] para el estudio de secuencias de ADN. En este caso el criterio ordena los atributos en función de la pérdida de capacidad de predicción del sistema cuando un atributo es eliminado. En [14] los autores aproximan el rendimiento en generalización del sistema cuando el atributo  $i$ -ésimo es eliminado, por el acierto en reescritura cuando se sustituye el valor de ese atributo en todos los ejemplos por su media. Cuando se usa este método para el aprendizaje de preferencias, el criterio de ordenación puede representarse como

$$R_2(i) = \left( \sum_k z_k \cdot \sum_j \alpha_j z_j \mathcal{K}(\mathbf{x}_j^{(1),i}, \mathbf{x}_j^{(2),i}, \mathbf{x}_k^{(1),i}, \mathbf{x}_k^{(2),i}) \right) \quad (9)$$

donde  $x^i$  denota a un vector en el que se ha sustituido el valor del  $i$ -ésimo atributo por su media. Nótese que un valor alto de  $R_2(i)$ , esto es, una tasa de acierto alta sobre el conjunto de entrenamiento cuando se modifica el valor del atributo  $i$ -ésimo, equivale a una menor relevancia de ese atributo. De esta manera, el atributo a eliminar será el que maximice el anterior criterio, al contrario que en el método anterior.

## 4.2. Selección del Modelo

A partir de una ordenación de atributos se obtiene una secuencia de subconjuntos anidados, cada uno de ellos resultante de eliminar un atributo del conjunto anterior. A partir de la lista ordenada de subconjuntos de atributos, el siguiente paso es seleccionar uno de entre todos los modelos que se inducen a partir de aquellos. En general nuestro interés estará centrado en obtener un subconjunto que haga que el sistema de aprendizaje alcance el máximo acierto en generalización posible.

La alternativa seleccionada es un método basado en métricas denominado ADJ (ADJusted distance estimate) [15,16]. Este método está diseñado para seleccionar el nivel de complejidad apropiado para ajustarse a los datos existentes. En este caso, y dada la secuencia anidada de subconjuntos de atributos devuelta por los métodos presentados anteriormente,  $F_1 \subset F_2 \subset \dots \subset F_n$ , ADJ nos proveerá de un procedimiento para seleccionar uno de los hiperplanos  $g_i$ , inducidos por SVM a partir de sus respectivos subconjuntos  $F_i$ .

La idea central de ADJ es la definición de una métrica en el espacio de hipótesis. Así, para dos hipótesis  $f$  y  $g$ , su distancia se puede calcular como el desacuerdo o discrepancia entre sus predicciones

$$d(f, g) \stackrel{def}{=} \varphi \left( \int err(f(x), g(x)) dP_X \right) \quad (10)$$

donde  $err(f(x), g(x))$  es una medida de la discrepancia en un punto cualquiera  $x$  en el espacio  $X$  de descripciones de ejemplos. Dado que estas distancias sólo pueden ser aproximadas, ADJ define un método para calcular  $\hat{d}(g, t)$ , o *estimación ajustada de distancia* entre una hipótesis cualquiera  $g$  y la función *real* de clasificación  $t$ . Así, la hipótesis seleccionada será

$$g_k = \arg \min_{g_l} \hat{d}(g_l, t). \quad (11)$$

La estimación de la distancia,  $\hat{d}$ , se calcula a partir de las discrepancias esperadas en las predicciones sobre un par de conjuntos: el conjunto de entrenamiento  $T$ , y un conjunto  $U$  de ejemplos sin etiquetar, esto es, un conjunto de ejemplos muestreados a partir de la distribución de probabilidad *real* del problema  $P_X$ , pero para los que se desconoce su categoría. La estimación de ADJ se define como

$$ADJ(g_l, t) \stackrel{def}{=} d_T(g_l, t) \cdot \max_{k < l} \frac{d_U(g_k, g_l)}{d_T(g_k, g_l)} \quad (12)$$

donde, para un determinado subconjunto de ejemplos  $S$ ,  $d_S(f, g)$  se define como la discrepancia esperada de las hipótesis  $f$  y  $g$  sobre el conjunto  $S$ . Aquí se propone el uso de la corrección de Laplace a la estimación de la probabilidad de la siguiente manera:

$$d_S(f, g) \stackrel{def}{=} \frac{1}{|S| + 2} \left( 1 + \sum_{i \in S} 1_{f(x_i) \neq g(x_i)} \right) \quad (13)$$

Nótese que, mediante esta modificación, se puede discriminar entre dos hipótesis cuya discrepancia sobre el conjunto  $T$  sea cero (Véase la Ecuación 12).

Obtener un conjunto de ejemplos sin etiquetar no es una tarea trivial. En [17] los autores proponen un método de muestreo sobre el conjunto de entrenamiento para resolver este problema. Sin embargo, cuando se trata de aprender preferencias, se puede construir fácilmente un conjunto de ejemplos sin etiquetar a partir de juicios de preferencias formados por parejas de objetos reales seleccionados aleatoriamente entre los presentes en el conjunto de entrenamiento original. Por definición, se fijará el tamaño de  $U$  como 10 veces mayor que el de  $T$ .

### 4.3. Redundancia

En el ámbito del análisis de datos sensoriales no es raro que varios expertos tengan opiniones similares sobre una determinada característica. Algunas variables físicas y químicas pueden presentar también este comportamiento. Es por ello que este tipo de conjuntos pueden presentar frecuentemente un cierto grado de redundancia. Para intentar aprovechar este tipo de redundancias, se ha desarrollado un sistema, RF, simple aunque bastante efectivo, con el fin de ser aplicado al conjunto de entrenamiento antes de cualquier sistema de selección de atributos. Por otro lado, este sistema proporciona un beneficio adicional, al reducir la dimensión del problema que el método de selección de atributos elegido va a tener como entrada.

RF es un proceso iterativo que, en cada etapa, genera una nueva descripción a partir del conjunto de datos inicial. Los dos atributos más *similares* son reemplazados por uno nuevo, cuyos valores se calculan como la media de los anteriores. Dados dos atributos  $a_i$  y  $a_j$  como vectores (columna) cuya dimensión es el número de ejemplos en el conjunto de entrenamiento, la similitud puede ser estimada por medio del coseno del ángulo que forman, esto es,

$$\text{similitud}(a_i, a_j) = \frac{\langle a_i, a_j \rangle}{\|a_i\| \cdot \|a_j\|} \quad (14)$$

Aplicando este método se obtiene una secuencia de diferentes descripciones del conjunto inicial, cada una de ellas con un atributo menos que la anterior. Para seleccionar una descripción adecuada, en función de la capacidad de predicción, se usa de nuevo ADJ. La descripción seleccionada puede considerarse un resumen del conjunto de datos original, listo para ser procesado mediante los algoritmos de selección de subconjuntos de atributos descritos anteriormente.

## 5. Resultados Experimentales

Para ilustrar los beneficios obtenidos al utilizar las aproximaciones descritas, se han realizado una serie de experimentos con un par de bases de datos sensoriales. En los dos casos, se realizó una comparación entre las puntuaciones obtenidas por las aproximaciones basadas en preferencias y aquellas obtenidas por los métodos de regresión. En todos los casos, a la hora de estimar los errores, se ha usado una validación cruzada de 10 folds y 5 repeticiones.

Como ha sido explicado anteriormente, el concepto más importante es el de las sesiones de prueba. Así, en cada sesión, para resumir las opiniones de los consumidores, se ha calculado la media de las valoraciones obtenidas para cada producto. Esto da lugar a un conjunto de regresión, esto es, descripciones de objetos acompañados de una categoría continua. De forma adicional se pueden obtener algunos juicios de preferencias. El conjunto de regresión así construido puede usarse para inducir una función que predice las valoraciones de los consumidores. Estos experimentos se han realizado usando una regresión lineal simple, y un método de fama reconocida: Cubist, un producto comercial de la compañía RuleQuest Research [18].

Para interpretar los resultados se ha usado la desviación absoluta media relativa ( $rmad$ ). Este valor se calcula a partir de la desviación absoluta media ( $mad$ ) de la función  $f$  inducida por el método de regresión, mediante la siguiente expresión

$$mad = \frac{1}{|E'|} \sum (|f(x) - x_{clase}| : x \in E') \quad (15)$$

donde  $E'$  es un conjunto de test. El estimador  $rmad$  se calcula entonces como 100 veces el cociente de  $mad$  entre el  $mad$  del predictor constante e incondicional que devuelve el valor medio en todos los casos. Así,

$$rmad = 100 \cdot \frac{mad(f)}{mad(media)} \quad (16)$$

Nótese que una  $rmad$  del 100% significa que el método de regresión tiene la misma desviación absoluta media que el predictor constante de la media.

Por otra parte, cuando se trabaja con juicios de preferencias, los errores tienen una interpretación más clara. El error cometido es el número de juicios de preferencias mal clasificados. Para manejar estos conjuntos de ejemplos, se ha usado SVM con kernel lineal ( $SVM_l$ ) y polinómico de grado 2 ( $SVM_p$ ). Se le ha dado otra oportunidad a los métodos de regresión, entrenándolos con sus conjuntos, y probándolos con conjuntos de juicios de preferencias obtenidos a partir de pares de ejemplos de test ordenados de acuerdo a su categoría. Para cada ejemplo se han seleccionado aleatoriamente otros 10, para formar así 10 juicios de preferencias.

La primera base de datos es resultado de un estudio realizado con el fin de determinar qué atributos son más importantes de cara a la aceptación por parte del consumidor de la carne de vacuno [19]. El objetivo del estudio era probar la influencia de la raza del ganado y el tiempo de maduración de las piezas de



**Tabla 1.** Resultados sobre los conjuntos de la carne de vacuno y la sidra asturiana. Para los métodos de regresión se reseña la desviación absoluta media relativa del error. Para el aprendizaje de preferencias, se muestra el porcentaje de juicios de preferencias mal clasificados. El número de atributos seleccionado se muestra para los algoritmos FSS . Los conjuntos originales tienen 64 (sidra) y 147 (carne) atributos. Los resultados han sido obtenidos por medio de una validación cruzada de 10 folds.

	Regresión						Preferencias						Preferencias (SVM <sub>p</sub> +FSS )						
	Lineal		Cubist		SVM <sub>I</sub>		SVM <sub>p</sub>		FSS <sub>1</sub>		FSS <sub>2</sub>		RF		RF+FSS <sub>1</sub>		RF+FSS <sub>2</sub>		
	Rmad	Error	Rmad	Error	Error	Error	Error	Error	Error	#Atrs.	Error	#Atrs.	Error	#Atrs.	Error	#Atrs.	Error	#Atrs.	Error
acidez	103.0 %	109.4 %	40.0 %	42.4 %	29.9 %	18.0 %	17.2 %	18.2 %	17.6 %	16.9 %	16.4 %	47.5 %	14.7 %	19.3 %	19.3 %	18.5 %	18.5 %	22.9 %	27.6 %
amargor	105.8 %	111.9 %	56.0 %	47.4 %	30.5 %	23.1 %	25.1 %	29.4 %	21.2 %	22.0 %	20.8 %	39.0 %	18.2 %	23.9 %	23.9 %	22.9 %	27.6 %	22.9 %	27.6 %
sabor-1	105.3 %	111.7 %	42.4 %	44.3 %	27.2 %	17.1 %	18.5 %	30.0 %	21.4 %	26.2 %	20.6 %	36.0 %	20.1 %	24.1 %	24.1 %	19.7 %	22.2 %	19.7 %	22.2 %
sabor-2	107.2 %	116.0 %	45.6 %	45.0 %	28.6 %	17.9 %	19.1 %	27.0 %	15.6 %	26.3 %	17.8 %	40.5 %	17.8 %	22.9 %	22.9 %	16.5 %	18.1 %	16.5 %	18.1 %
sabor-3	110.3 %	107.7 %	43.8 %	41.8 %	33.6 %	17.7 %	22.7 %	28.8 %	21.8 %	19.8 %	19.3 %	30.0 %	20.1 %	17.6 %	18.4 %	18.3 %	18.3 %	18.4 %	18.3 %
bouquet	104.0 %	110.2 %	43.5 %	42.7 %	26.4 %	21.0 %	16.7 %	30.0 %	18.9 %	28.1 %	19.8 %	45.0 %	18.8 %	24.1 %	18.0 %	23.1 %	18.0 %	23.1 %	23.1 %
color	98.4 %	109.9 %	41.3 %	43.4 %	26.1 %	17.8 %	19.5 %	32.0 %	22.0 %	22.6 %	19.5 %	41.5 %	21.6 %	24.3 %	24.9 %	21.0 %	24.9 %	21.0 %	21.0 %
visual-1	103.2 %	113.0 %	41.7 %	43.1 %	25.9 %	13.4 %	11.5 %	30.4 %	13.8 %	24.5 %	12.0 %	34.5 %	14.7 %	24.5 %	13.8 %	18.1 %	13.8 %	18.1 %	18.1 %
visual-2	102.3 %	112.0 %	43.8 %	45.7 %	34.0 %	20.0 %	21.1 %	30.5 %	18.9 %	23.8 %	19.9 %	35.5 %	21.2 %	27.1 %	19.4 %	23.2 %	19.4 %	23.2 %	23.2 %
visual-3	107.2 %	120.5 %	45.6 %	49.3 %	25.3 %	20.6 %	16.1 %	18.4 %	15.6 %	18.0 %	13.8 %	32.5 %	13.4 %	24.6 %	13.4 %	25.3 %	13.4 %	25.3 %	25.3 %
visual-4	98.7 %	97.2 %	36.5 %	38.2 %	23.0 %	14.0 %	14.1 %	25.5 %	15.0 %	19.4 %	14.1 %	38.5 %	14.7 %	22.4 %	12.1 %	19.2 %	14.7 %	19.2 %	19.2 %
Promedio sidra	104.1 %	110.9 %	43.7 %	43.9 %	28.2 %	18.2 %	18.3 %	27.3 %	18.3 %	22.5 %	17.6 %	38.2 %	17.7 %	23.2 %	18.0 %	21.3 %	17.6 %	23.2 %	21.3 %
terneza	96.3 %	97.8 %	41.5 %	43.1 %	29.6 %	19.4 %	-	-	-	-	20.0 %	50.0 %	21.8 %	27.0 %	21.3 %	37.5 %	21.8 %	27.0 %	37.5 %
sabor	99.3 %	103.4 %	43.8 %	46.5 %	32.7 %	23.8 %	-	-	-	-	25.0 %	65.0 %	26.5 %	33.5 %	26.1 %	29.0 %	26.5 %	33.5 %	29.0 %
aceptación	94.0 %	97.2 %	38.4 %	40.2 %	31.9 %	22.1 %	-	-	-	-	24.7 %	39.5 %	24.8 %	30.0 %	25.3 %	26.7 %	24.8 %	30.0 %	26.7 %
Promedio carne	96.5 %	99.5 %	41.2 %	43.3 %	31.4 %	21.8 %	-	-	-	-	23.2 %	51.5 %	24.4 %	30.2 %	24.2 %	31.1 %	24.4 %	30.2 %	24.2 %
Promedio total	102.5 %	108.4 %	43.1 %	43.8 %	28.9 %	19.0 %	-	-	-	-	18.8 %	41.1 %	19.2 %	24.7 %	19.4 %	23.4 %	19.2 %	24.7 %	23.4 %

carne después del sacrificio del animal. Para conseguir esto se usó un conjunto de siete razas españolas. Cada consumidor valoró 4 o 5 piezas de carne en cada sesión. Las piezas de carne están descritas por: 12 atributos determinados por 11 diferentes expertos, el peso del animal, el tiempo de maduración y 6 medidas físicas que describen la textura. Dado que la raza está representada mediante 7 atributos booleanos, la descripción total de una pieza de carne usa 147 variables.

### 5.1. Preferencias frente a regresión

En primer lugar se ha realizado una comparación entre las aproximaciones basadas en preferencias y aquellas basadas en métodos de regresión. Para tener en cuenta el concepto de *sesión de pruebas*, se ha calculado la media de las puntuaciones obtenidas por un determinado producto en cada sesión. Estos conjuntos pueden ser usados para inducir una función que otorga valoraciones a los productos en función de los gustos de los consumidores. También aquí se han usado un método de regresión lineal básico y Cubist, una potente herramienta comercial. Por otro lado, se han construido a su vez los conjuntos de preferencias en la forma acostumbrada.

La Tabla 1 muestra los resultados conseguidos por todos los métodos anteriores usando una validación cruzada de 10 folds. Se han incluido también los resultados del capítulo anterior, con el fin de facilitar una mejor comparación. Los resultados muestran que los métodos basados en regresión son incapaces de aprender. Sus desviaciones absolutas medias relativas (*rmad*) son cercanas al 100%. Los métodos basados en preferencias obtienen mejores resultados, sobre todo aquellos que utilizan kernels no lineales.

### 5.2. Selección de atributos

Para encontrar aquellos atributos que tienen mayor influencia en los gustos de los consumidores, se han aplicado los métodos de selección de atributos descritos en la sección anterior. Por simplicidad, se denotará como  $FSS_1$  y  $FSS_2$  a los métodos de selección de atributos 1 y 2 respectivamente. Además, se ha usado RF de dos maneras distintas: como seleccionador de un subconjunto de atributos, y como un filtro aplicable antes de  $FSS_1$  y  $FSS_2$ . En todos los casos se ha usado ADJ para seleccionar el modelo más apropiado. El sistema de aprendizaje usado ha sido  $SVM_p$ , dado que fue el más preciso en las anteriores pruebas. En los conjuntos de la carne es imposible el uso directo de los sistemas FSS, debido su alto coste computacional, si no se consigue antes una reducción en el número de atributos. Por lo tanto, en la tabla sólo se presentan los resultados sobre esos conjuntos cuando se ha aplicado RF como filtro previo. Lo que es más, en esos conjuntos los atributos fueron eliminados en lotes de 5, con el fin de poder obtener los resultados en un tiempo aceptable. En los conjuntos de la sidra, de menores dimensiones, no se aplicó ninguna de estas restricciones, esto es, los atributos fueron eliminados de uno en uno, y los métodos de selección de atributos son aplicables de forma autónoma.

Se puede ver (Tabla 1) que  $FSS_1$ ,  $FSS_2$  y RF reducen considerablemente el número de atributos sin pérdida de precisión en la mayoría de los casos. En los conjuntos de la sidra, todos los métodos obtienen resultados similares (no existen diferencias significativas), pero  $FSS_2$  es significativamente mejor que  $FSS_1$  cuando se trata de reducir el problema, mientras que RF es el que peores resultados consigue. En estos mismos conjuntos, los porcentajes de acierto mejoran ligeramente cuando se usa RF como filtro previo. Sin embargo, en los conjuntos de la carne, el acierto decrece ligeramente cuando se usa RF con respecto a lo obtenido por  $SVM_p$ . De hecho, también decrece cuando se usan  $FSS_1$  y  $FSS_2$  después de RF. Este comportamiento puede ser debido a la necesidad de eliminar los atributos en lotes de cinco por culpa de las excesivas dimensiones del problema.

## 6. Conclusiones

Se ha mencionado anteriormente hasta qué punto es importante el intentar descubrir la información relevante en los procesos de análisis de la calidad sensorial. Estos conjuntos presentan unas características que los hacen intratables para métodos clásicos de regresión. Aparte, también está presente el *efecto lote*, que desvirtúa las puntuaciones otorgadas por los expertos.

El aprendizaje de preferencias no intenta obtener las valoraciones exactas. Sin embargo, es capaz de encontrar modelos que son capaces de explicar los gustos de los consumidores. Se ha observado a su vez que la capacidad de acierto se ve notablemente mejorada cuando se usan métodos no lineales sobre las dos bases de datos analizadas. En general, la utilidad de estos modelos puede verse incrementada mediante el uso de métodos de selección de atributos especialmente preparados.

Otro interesante aspecto de los conjuntos de datos sensoriales es que, frecuentemente, existen bloques de atributos que describen el mismo aspecto del producto. Para sacar provecho de estas redundancias se ha desarrollado un pequeño filtro que puede ser aplicado antes de la selección de los atributos para mejorar la eficacia del sistema de aprendizaje.

## Referencias

1. Næs, T., Risvik, E.: Multivariate analysis of data in sensory science. Elsevier (1996)
2. Murray, J., Delahunty, C., Baxter, I.: Descriptive sensory analysis: past, present and future. *Food Research International* **3** (2001) 461–471
3. Buck, D., Wakeling, I., Greenhoff, K., Hasted, A.: Predicting paired preferences from sensory data. *Food quality and preference* **12** (2001) 481–487
4. Corney, D.: Designing food with bayesian belief networks. En: *Proceedings of the International Conference on Adaptative Computing in engineering Design and Manufacture*. (2002) 83–94

5. Herbrich, R., Graepel, T., Obermayer, K.: Support vector learning for ordinal regression. En: Proceedings of the Ninth International Conference on Artificial Neural Networks, Edimburgh,UK (1999) 97–102
6. Joachims, T.: Optimizing search engines using clickthrough data. En: Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD). (2002)
7. Díez, J., del Coz, J., Luaces, O., Goyache, F., Alonso, J., Peña, A., Bahamonde, A.: Learning to assess from pair-wise comparisons. In Springer-Verlag, ed.: Proceedings of the 8th Iberoamerican Conference on Artificial Intelligence, Sevilla,Spain (2002) 481–490
8. Cohen, W., Shapire, R., Singer, Y.: Learning to order things. *Journal of Artificial Intelligence Research* **10** (1999) 243–270
9. Dumais, S., Bharat, K., Joachims, T., Weigend, A., eds.: Workshop on implicit measures of user interests and preferences. In ACM SIGIR Conference, Toronto, Canada (2003)
10. Goyache, F., A.Bahamonde, Alonso, J., López, S., del Coz, J., Quevedo, J., Ranilla, J., Luaces, O., Álvarez, I., Royo, L., Díez, J.: The usefulness of artificial intelligence techniques to assess subjective quality of products in the food industry. *Trends in Food Science and Technology* **12** (2001) 370–381
11. Herbrich, R., Graepel, T., Obermayer, K.: 7. En: Large Margin Rank Boundaries for Ordinal Regression. MIT Press, Cambridge, MA (2000) 97–102
12. Guyon, I., Weston, J., Barnhill, S., V.Vapnik: Gene selection for cancer classification using support vector machines. *Machine Learning* **46** (2002) 389–422
13. Rakotomamonjy, A.: Variable selection using svm-based criteria. *Journal of Machine Learning Research* **3** (2003) 1357–1370
14. Degroeve, S., Baets, B.D., de Peer, Y.V., Rouzé, P.: Feature subset selection for splice site prediction. *Bioinformatics* **18** (2002) 75–83
15. Schuurmans, D.: A new metric-based approach to model selection. *AAAI/IAAI* (1997) 552–558
16. Schuurmans, D., Southey, F.: Metric-based methods for adaptive model selection and regularization. *Machine Learning* **48** (2002) 51–84
17. Bengio, Y., Chapados, N.: Extensions to metric-based model selection. *Journal of Machine Learning Research* **3** (2003) 1209–1227
18. RuleQuest: Computer software for regression. RuleQuest Research. (2001)
19. Gil, M., Serra, X., Gispert, M., Oliver, M., Sañudo, C., Panea, B., Olleta, J., Campo, M., Oliván, M., Osoro, K., García-Cachan, M., Cruz-Sagredo, R., Izquierdo, M., Espejo, M., Martín, M., Piedrafitra, J.: The effect of breed-production systems on the myosin heavy-chain 1, the biochemical characteristics and the colour variables of longissimus thoracis from seven spanish beef cattle breeds. *Meat Science* **58** (2001) 181–188
20. Díez, J., Bayón, G., Quevedo, J., del Coz, J., Luaces, O., Alonso, J., Bahamonde, A.: Discovering relevancies in very difficult regression problems: applications to sensory data analysis. En: Proceedings of the European Conference on Artificial Intelligence (Aceptado para publicación). (2004)
21. Luaces, O., Bayón, G., Quevedo, J., Díez, J., del Coz, J., Bahamonde, A.: Analyzing sensory data using non-linear preference learning with feature subset selection. En: Proceedings of the European Conference on Machine Learning (Aceptado para publicación), Pisa, Italy (2004)
22. Bahamonde, A., Bayón, G., Díez, J., Quevedo, J., Luaces, O., del Coz, J., Alonso, J.: Feature subset selection for learning preferences: a case study. En: Proceedings of the International Conference on Machine Learning (Aceptado para publicación), Banff, Alberta, Canada (2004)