

# Clustering people according to their preference criteria

Jorge Díez\*, Juan José del Coz, Oscar Luaces, and Antonio Bahamonde

Centro de Inteligencia Artificial. Universidad de Oviedo at Gijón,

Campus de Viesques, E-33271 Gijón (Asturias), Spain

{jdiez, juanjo, oluaces, antonio}@aic.uniovi.es

**Abstract.** Learning preferences is a useful task in application fields such as collaborative filtering, information retrieval, adaptive assistants or analysis of sensory data provided by panels. SVMs, using preference judgments, can induce ranking functions that map objects into real numbers, in such a way that more preferable objects achieve higher values. In this paper we present a new algorithm to build clusters of people with closely related tastes, and hence people whose preference judgment sets can be merged in order to learn more reliable ranking functions. In some application fields, these clusters can be seen as market segments that demand different kinds of products. The method proposed starts representing people's preferences in a metric space, where it is possible to define a kernel based similarity function; finally a clustering algorithm discovers significant groups with homogeneous tastes. The key point of our proposal is to use the ranking functions induced from the preference judgments of each person; we will show that those functions codify the criteria used by each person to decide her preferences. To illustrate the performance of our approach, we present two experimental cases. The first one deals with the collaborative filtering database EachMovie. The second database describes a real case of consumers of beef meat.

**Keywords:** learning preferences, clustering, adaptive assistants, analysis of sensory data, market segmentation.

---

\* Corresponding author. *Fax:* +34 985 18 21 25. *Phone:* +34 985 18 25 88. *E-mail address:* jdiez@aic.uniovi.es

## 1. Introduction

Supervised inductive learning deals with sets of training examples; these represent pairs of input and the attached outputs of a function that has to be found in a given family of hypotheses. The input is described by a set of attribute values, while the output is in fact another attribute of the examples called class; its type determines the approach and even the name of the learning task. Regression is used when the class is a continuous number, and categorical classification is employed when the class or output of training examples is one of a finite set of symbolic categories.

In this paper we tackle a slightly different problem: learning people's preferences for consumable products, or for system configurations, or for responding to information requests. Here the training material can be expressed as in regression problems: the description of each object is then followed by a number that assesses the degree of satisfaction. Alternatively, training examples can be represented by *preference judgments*: pairs of vectors  $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$  where someone expresses the fact that he or she prefers  $\mathbf{x}^{(1)}$  to  $\mathbf{x}^{(2)}$ . In other words, training sets are samples of binary relations between objects.

As pointed out in (Cohen et al., 1999; Dumais et al., 2003), obtaining preference information may be easier and more natural than obtaining the labels needed for a classification or regression approach. Moreover, this type of information is more accurate, since people tend to rate their preferences in a relative way, comparing objects with the other partners in the same batch. There is a kind of *batch effect* that often biases the ratings. Thus, an object presented in a batch surrounded by worse objects will probably obtain a higher rating than if it were presented together with better objects.

There are algorithms in the literature able to learn preferences. Sometimes they aim to classify pairs of objects  $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ , deciding whether  $\mathbf{x}^{(1)}$  is preferable to  $\mathbf{x}^{(2)}$  or not, as in (Branting and Broos, 1997; Cohen et al., 1999). Another approach consists in learning a real *preference* or *ranking function*  $f$  from the space of objects considered in such a way that  $f(\mathbf{x}^{(1)}) > f(\mathbf{x}^{(2)})$  whenever  $\mathbf{x}^{(1)}$  is preferable to  $\mathbf{x}^{(2)}$ . This functional approach can start from a set of objects endowed with a (usually ordinal) rating, as in regression (Herbrich et al., 2000; Crammer and Singer, 2001; Shashua and Levin, 2002), or can stem from sets of preference judgments, as in (Tesauro, 1989; Utgoff and Clouse, 1991; Freund et al., 1998; Fiechter and Rogers, 2000; Joachims, 2002; Bahamonde et al. 2004).

In this paper we present a new algorithm that, given a family of preference judgment sets from a number of people, discovers groups with homogeneous tastes. The usefulness of this task is twofold. First, we can merge the sets of preference judgments of members of the same group, thus attaining more useful and reliable knowledge about peoples' preferences; this is usually a critical point, since the available data from individuals is frequently quite limited.

In second place, the discovery of clusters is important by itself. This is the case when we deal with data collected from panels of consumers; then the groups found may constitute significant market segments that demand different kinds of products. At the end of the paper, we will present an application to food industry. Using the data from a panel of beef meat consumers, we discuss the implications of the results returned by the algorithm for meat industries and breeders. A shorter version of this case study was presented in (Díez et al. 2005).

To cluster anything it is necessary to define a similarity measure. The core idea of our proposal is that similarity can be defined between the ranking functions learned from the preference judgment set of each person; these functions codify the *criteria* used to decide the preferences. Since those ranking functions are induced by a classification Support Vector Machine (SVM) (Vapnik, 1998), the similarity measure is defined by means of a kernel method.

In the next section, we review the literature related to explaining the usefulness of clusters of preference criteria in different areas of application. Then, we discuss the details of our proposal. Finally, the paper concludes with a report of the experiments conducted to evaluate the proposed algorithm. For this purpose we use EachMovie (McJones, 1997), a publicly available collaborative filtering database for movie ratings. Finally, we review the results obtained with the dataset of the panel of beef meat consumers.

## **2. How clusters can be useful**

The learning tasks involved in *recommender* systems (Resnick and Varian, 1997) can be considered as special cases of ordinal regression. Here users rate one kind of object and receive recommendations about

objects that they are likely to prefer. Such advice can be elaborated according to the relationship of the properties of the objects and the user's past ratings; this is the *content-based* model (Basu et al., 1998; Pazzani, 1999). Or, on the other hand, in the model called *collaborative* or *social filtering*, the recommendations are induced from the user and other users' ratings, formulating them as a learning task (Goldberg et al., 1992; Resnick et al., 1994; Shardanand and Maes, 1995).

Within this context, as pointed out in (Hofmann and Puzicha, 1999), there is another fundamental problem in addition to the prediction of ratings: the discovery of meaningful groups or *clusters* of persons and objects able to explain the observed preferences by some smaller number of typical preference patterns. This point of view gives rise to the *latent class* model (Cheung et al., 2000). See also (Ungar and Foster, 1998).

There are other application fields where clusters and preferences appear together as a desirable mixture. For instance, in (Joachims, 2002), Joachims presents an information retrieval system equipped with a ranking function learned from click-through data collected from user interaction with a www search engine. To improve his proposal, the author acknowledges the need to obtain feasible training data. This raises the question of the convenience of developing clustering algorithms to find homogeneous groups of users.

An Adaptive Route Advisor is described in (Fiechter and Rogers, 2000); the system is able to recommend a route to lead users through a digitalized road map taking into account their preferences. An interesting extension discussed in the paper is to modify route recommendations depending on the time of the day or the purpose of the trip. The approach suggested the inclusion of an algorithm that clusters user preferences into contexts.

### **3. Sensory data Analysis**

The quality or acceptability of market products can be measured in a number of different dimensions. Sensory analysis is concerned with those aspects that are principally appreciated through sensory impressions. It is typically used by food industries and breeders to improve some production decisions.

An excellent survey of the use of this type of data in the food industry can be found in (Murray et al., 2001; Buck et al., 2001); for a Machine Learning perspective, see (Corney, 2002) and (Goyache et al., 2001; Del Coz et al. 2004; Luaces et al. 2004; Díez et al. 2005).

From a conceptual point of view, sensory data include the assessment of products provided by two different kinds of panels. The first one is made up of a small group of expert, trained judges; these will describe each product by attribute-value pairs. Expert panelists are thus required to have enough sensory accuracy so as to discriminate between different and similar products; note that experts are not necessarily asked to rate the overall quality of products. This panel will play the role of a bundle of sophisticated sensors. So, experts' assessments are used to describe the products in addition to other measures usually obtained by some chemical, biological or physical devices. The second kind of panel is made up of untrained consumers; these are asked to rate their degree of acceptance of the tested products on a scale.

The aim of the analysis of these data is to be able to relate product descriptions with consumer preferences. Therefore, sensory studies of markets will start out from tables such as Table 1. Each row represents a product rated by a consumer in a given tasting *session*. The concept of session is important, since we can interpret consumer ratings relative to each session (Joachims, 2002; Bahamonde et al. 2004; Díez et al. 2004). In this way, we do not need to assume that a rating of “7” means the same thing to every consumer and in every session (Cohen et al., 1999).

Additionally, when dealing with food products, we must realize that the size of the sample prevents panelist from testing all products. Hence, we cannot ask our panelist to spend long periods rating the whole set of food samples. Typically, each consumer only participates in one or a small number of tasting sessions, usually in the same day. Notice that tasting a large sample of food may be physically impossible, or the number of tests performed would damage the sensory capacity of consumers.

On the other hand, expert descriptions are ratings in an ordinal scale of different aspects of products related to their taste, odor, color, etc. Here we must assume that a rating of “7” (in say, texture) means the same for a given expert in every product; though not necessarily for every expert.

Sometimes, to describe a kind of food products, it is possible to use categorical attributes instead of numerical ones. In these cases, it will be necessary the use of an adequate kernel in the algorithms that we are going to explain in the next section.

From a practical point of view, the market is not interested in tastes of individual consumers, the purpose of marketing studies of sensorial data is to discover, if they exist, widespread ways to appreciate food products that can be considered as *market segments*. These segments can be seen as clusters of consumers with similar tastes. In this paper, we will show that the similarity of preference criteria of consumers can be computed in a high dimension space by means of a kernel-based method.

#### 4. Clustering people according to their preferences

In this section we are going to discuss the available options to build a method for clustering people using their preference criterion as the measure of similarity. For ease of reference, let  $S$  be a set of  $n$  objects from an input space  $X$  that will be rated by  $m$  people. Therefore, given a space of ordinal values,  $Scale$ , we have a family

$$r_i: S_i \rightarrow Scale, \quad i = 1, \dots, m \quad (1)$$

of rankings, one for each person, that in general are partially defined with respect to  $S$ , since usually not everybody rates everything, that is,  $S_i \subseteq S$ .

To measure the similarity between the preferences of two people  $p_i$  and  $p_j$ , a first attempt consists of comparing the vectors  $(r_i(\mathbf{x}): \mathbf{x} \in S_i)$  and  $(r_j(\mathbf{x}): \mathbf{x} \in S_j)$  of their ratings. To do this, we must realize that  $S_i$  and  $S_j$  can have few common objects. In fact, in sensory data it is frequent that  $S_i \cap S_j = \emptyset$ .

Different tools have been employed to make comparisons when using these rating vectors for prediction tasks in collaborative filtering (see (Breese et al., 1998)); the most obvious and unsuccessful being Euclidean distance, used in nearest neighbor algorithms. Pearson's correlation or the cosine of the vectors has been put forward to consider the possible differences in the scales used by different people.

However, these comparison techniques, devised for prediction purposes, are not easily extendable to clustering. To illustrate this point, let us consider one person  $p_i$  with a coherent preference criterion, and let us divide  $p_i$ 's rating vector in two parts:

$$(r_i(\mathbf{x}): \mathbf{x} \in S_{i1}), (r_i(\mathbf{x}): \mathbf{x} \in S_{i2}), \text{ with } S_{i1} \cap S_{i2} = \emptyset \text{ and } S_{i1} \cup S_{i2} = S_i. \quad (2)$$

These two vectors would not have anything in common for any reasonable comparison measure. However, both vectors represent the same *rating criterion*, the criterion of  $p_i$ ; so  $p_i$  rating both  $S_{i1}$  and  $S_{i2}$  must be included in the same cluster.

However, it is not obvious how to represent what we have lazily called rating criterion of a person. The proposal presented in this paper is to represent preference criteria explicitly by means of a function able to generalize somehow the preferences provided by the ratings. The following sections will show how these functions can be defined and learned. Then we will describe the clustering method based on the similarity of these functions.

But before going on it is necessary to reconsider the kind of data that we will use as training material to learn people's preferences. If we have a vectorial way to describe the objects in  $S$  and a rating vector  $(r(\mathbf{x}): \mathbf{x} \in S)$ , we can try to use regression to induce a function that maps object descriptions into ratings. However, this is not a faithful way of capturing people's preferences; see (Herbrich et al., 2000) for a discussion of the differences between *ordinal* and *continuous* (metric or least squares) regression.

Additionally, when consumer panels are involved, an important reason to discard any kind of regression is that ratings are relative orderings instead of absolute values, due to the so-called *batch effect* (Joachims, 2002; Díez et al., 2004). Therefore, when it is possible to consider explicitly the tasting session where the rates were made, we will take them into account. Thus, a more general setting will be used. Following (Herbrich et al., 2000; Joachims, 2002), in order to capture the ordinal nature of ratings, and the relative character of people's preferences, for each consumer  $p_i$ , we will separate the ratings given at each tasting session. Then we will create a *preference judgment set*

$$PJ_i = \{(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) : \mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in S_i, r_i(\mathbf{x}^{(1)}) > r_i(\mathbf{x}^{(2)}), \text{ session}(p_i, \mathbf{x}^{(1)}) = \text{session}(p_i, \mathbf{x}^{(2)})\} \quad (3)$$

suitable for our needs just considering all pairs  $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$  of object in  $S_i$  such that they were presented in the same session to consumer  $p_i$ , and the rating of the first object was higher than the rating of the second.

#### 4.1 Class separation and preferences

Let us first consider a simplified case in this subsection before presenting the general learning framework. So, let us assume that the input space  $X$  is in fact a subset of real vectors; in symbols,  $X \subset \mathbf{R}^d$ . In order to learn our preference problems, we will try to find a real *ranking (or preference) function*  $f: \mathbf{R}^d \rightarrow \mathbf{R}$  that maximizes the probability of having  $f(\mathbf{x}^{(1)}) > f(\mathbf{x}^{(2)})$  whenever  $\mathbf{x}^{(1)}$  is preferable to  $\mathbf{x}^{(2)}$ . If we *now* restrict the space of hypothesis to linear functions, each preference judgments set  $PJ_i$  gives rise to a set of constraints

$$(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \in PJ_i \Leftrightarrow f_i(\mathbf{x}^{(1)}) > f_i(\mathbf{x}^{(2)}) \Leftrightarrow f_i(\mathbf{x}^{(1)} - \mathbf{x}^{(2)}) > 0 \Leftrightarrow f_i(\mathbf{x}^{(2)} - \mathbf{x}^{(1)}) < 0 \quad (4)$$

Therefore, ranking functions can be learned using a binary classification algorithm able to discriminate the class of objects according to the sign returned, as happens with Support Vector Machines (SVM). Notice that the training set is

$$T_i = \{(\mathbf{x}^{(1)} - \mathbf{x}^{(2)}; +1), (\mathbf{x}^{(2)} - \mathbf{x}^{(1)}; -1) : (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \in PJ_i\}. \quad (5)$$

The function learned passes through the origin of coordinates, and it is then defined by

$$f_i(\mathbf{x}) = \langle \mathbf{w}_i, \mathbf{x} \rangle = \sum_{k=1}^d \mathbf{w}_{ik} \mathbf{x}_k \quad (6)$$

where  $\mathbf{w}_i$  is the weight vector of the consumer  $p_i$ . The return of  $f_i$  on an object representation  $\mathbf{x}$  can be thought as the *assessment* of  $\mathbf{x}$  in the sense that  $f_i(\mathbf{x})$  will be used to predict preferences between  $\mathbf{x}$  and other products. On the other hand, the weight vector  $\mathbf{w}_i$  is the director vector of a hyperplane  $(\langle \mathbf{w}_i, \mathbf{x} \rangle = 0)$  called the *assessment hyperplane*. From a geometrical point of view, the distance (in the direction of  $\mathbf{w}_i$ ) from the hyperplane to each object is proportional to the value returned by  $f_i$ . In fact (see Figure 1),



$$\text{distance}(\langle \mathbf{w}_i, \mathbf{x} \rangle = 0; \mathbf{x}') = \frac{\langle \mathbf{w}_i, \mathbf{x}' \rangle}{\|\mathbf{w}_i\|} = \frac{f_i(\mathbf{x}')}{\|\mathbf{w}_i\|} \quad (7)$$

Of course, it would be difficult to find a linear function able to separate the positive from the negative differences of the training set  $T_i$ . In terms of preferences, not ever the assessment of an object will be proportional to the components of a vectorial description. It is not true that the more (or the less) the better; for instance, the amount of sugar or salt in our favorite foods usually has an optimal point and the increase or decrease from that equilibrium point is frequently rejected. Additionally, there may be situations where the objects considered have not a straightforward description by means of real vectors. To take into account all these issues, it is possible to use a *kernel trick* (Herbrich et al., 2000).

To present this general framework, let us assume that the space  $X$  of objects can be mapped to a Hilbert space  $\mathcal{F}$  using  $\phi: X \rightarrow \mathcal{F}$ . Then, as we have just seen, to learn a ranking or preference function we only have to find a *linear* separation in  $\mathcal{F}$  for the training set

$$T_{i\mathcal{F}} = \{(\phi(\mathbf{x}^{(1)}) - \phi(\mathbf{x}^{(2)}); +1), (\phi(\mathbf{x}^{(2)}) - \phi(\mathbf{x}^{(1)}); -1) : (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \in P_j\}. \quad (8)$$

To avoid handling the components of product descriptions in  $\mathcal{F}$ , the kernel trick helps us redefining  $T_i$  as

$$T_i = \{(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}; +1), (\mathbf{x}^{(2)}, \mathbf{x}^{(1)}; -1) : (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \in P_j\}. \quad (9)$$

Therefore, the input space that we need is in fact the product of  $X$  by itself, and it is mapped into the feature space  $\mathcal{F}$  using

$$\Psi: X \times X \rightarrow \mathcal{F}, \Psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \phi(\mathbf{x}^{(1)}) - \phi(\mathbf{x}^{(2)}). \quad (10)$$

Hence, the associated kernel to this transformation is

$$\begin{aligned} \mathbf{K}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}; \mathbf{x}^{(3)}, \mathbf{x}^{(4)}) &= \langle \Psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}), \Psi(\mathbf{x}^{(3)}, \mathbf{x}^{(4)}) \rangle = \langle (\phi(\mathbf{x}^{(1)}) - \phi(\mathbf{x}^{(2)})), (\phi(\mathbf{x}^{(3)}) - \phi(\mathbf{x}^{(4)})) \rangle \\ &= \langle \phi(\mathbf{x}^{(1)}), \phi(\mathbf{x}^{(3)}) \rangle - \langle \phi(\mathbf{x}^{(1)}), \phi(\mathbf{x}^{(4)}) \rangle - \langle \phi(\mathbf{x}^{(2)}), \phi(\mathbf{x}^{(3)}) \rangle + \langle \phi(\mathbf{x}^{(2)}), \phi(\mathbf{x}^{(4)}) \rangle \\ &= K(\mathbf{x}^{(1)}, \mathbf{x}^{(3)}) - K(\mathbf{x}^{(1)}, \mathbf{x}^{(4)}) - K(\mathbf{x}^{(2)}, \mathbf{x}^{(3)}) + K(\mathbf{x}^{(2)}, \mathbf{x}^{(4)}) \end{aligned} \quad (11)$$

where  $K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$  is the kernel associated to the transformation  $\phi$  of the *individual* objects. The kernel  $\mathbf{K}$  so built is called the Herbrich's Kernel (Herbrich et al., 2000) attached to  $K$ .

The separation function induced by a classification SVM from  $T_i$  with kernel  $K$  will be a function  $F_i: X \times X \rightarrow \mathbf{R}$  of the form

$$\mathbf{F}_i(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \sum_{s \in SV_i} \alpha_s y_s \langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}^{(1)}) - \phi(\mathbf{x}^{(2)}) \rangle = \sum_{s \in SV_i} \alpha_s y_s K(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \quad (12)$$

where  $SV_i$  is a set of indexes for the support vectors, and  $y_s$  is the class (+1 or -1) of the pairs  $(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)})$  of  $T_i$ . This function has an important property that is an immediate consequence of the kernel definition:

$$F_i(\mathbf{x}^{(1)}, 0) > F_i(\mathbf{x}^{(2)}, 0) \Leftrightarrow F_i(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) > 0 \quad (13)$$

Thus, we define the following assessment function

$$f_i: X \rightarrow \mathbf{R}, \quad f_i(\mathbf{x}) = F_i(\mathbf{x}, 0). \quad (14)$$

It is trivial to see that  $f_i$  is as coherent with the preference judgments of  $PJ_i$  as  $F_i$  is a separating function for  $T_i$ . Moreover, in the case of using as  $K$  the linear kernel,  $f_i$  coincides with the function defined in (6).

The expression of  $f_i$  can be simplified given that it is possible to skip a constant term from  $F_i(\mathbf{x}, 0)$ ; in the linear case this constant is 0. Therefore, in practice, the general assessment function  $f_i$  is given by

$$f_i(\mathbf{x}) = \sum_{s \in SV_i} \alpha_s y_s \langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}) \rangle = \sum_{s \in SV_i} \alpha_s y_s (K(\mathbf{x}_s^{(1)}, \mathbf{x}) - K(\mathbf{x}_s^{(2)}, \mathbf{x})) \quad (15)$$

## 4.2 Defining a similarity measure of people's preferences

Let us start with a motivating example assuming that there are 3 people that have been asked about their preferences about a kind of products. To simplify the presentation, let us suppose that the products studied can be represented by a single number. Think, for instance, in the proportion of milk to coffee in your *café au lait*. If the first person ( $p_1$ ) prefers coffees with a proportion of 0.20 of milk, her preference function could be a parabola with vertex in that point. Given that ranking or preference functions cross the origin of coordinates, the preference function of  $p_1$  has the form

$$f_1(x) = a_1 * (-x^2 + x * 2 * 0.20) = \langle a_1 * (-1, 0.40), (x^2, x) \rangle \quad (16)$$

where  $a_1$  is a positive constant. If the other two persons prefer the combinations around 0.21 and 0.40 respectively, the director vectors of the preference functions of each people are the following

$$\mathbf{w}_1 = a_1 * (-1, 0.40); \quad \mathbf{w}_2 = a_2 * (-1, 0.42); \quad \mathbf{w}_3 = a_3 * (-1, 0.80). \quad (17)$$

Notice that the values of positive constants  $a_i$  are not relevant if we want to use the functions  $f_i$  to predict the preference of  $p_i$  about a pair of possible combinations. What is really important is the relative weight of the components of  $\mathbf{w}_i$ . Thus, a way to measure the similarity of the tastes of these people is by the cosine of the director vectors of their preference functions

$$\text{similarity}(p_i, p_j) = \cos(\mathbf{w}_i, \mathbf{w}_j) = \frac{\langle \mathbf{w}_i, \mathbf{w}_j \rangle}{\|\mathbf{w}_i\| \cdot \|\mathbf{w}_j\|} \quad (18)$$

Using this definition, the similarity of  $p_1, p_2, p_3$  is

$$\text{similarity}(p_1, p_2) = 0.9999; \text{similarity}(p_1, p_3) = 0.9570; \text{similarity}(p_2, p_3) = 0.9618. \quad (19)$$

In other words, as one could hope, using this approach the tastes of  $p_1$  and  $p_2$  are nearer from each other than the taste of  $p_3$ .

Notice that we have summarized the tastes of people in their preference function instead of using somehow the set of ratings assigned to a collection of products. Probably the products tasted were different for each person, what would make difficult to define any reliable similarity measure based on these ratings.

Returning back to the general setting, let  $((P_j, \mathbf{w}_j): j = 1, \dots, m)$  be a collection of  $m$  preference judgment sets  $P_j$  endowed with the director vector  $\mathbf{w}_j$  of their respective ranking functions. We define the similarity of their preference criterion by the cosine of equation (18).

To make operative this definition, we need to arrange the equation (15) to make visible the director vector  $\mathbf{w}_i$  attached to a set of preference judgments  $P_j$ . In fact, the assessment function  $f_i$  can be represented by a weight vector  $\mathbf{w}_i$  in the higher dimensional space of features such that

$$f_i(\mathbf{x}) = \langle \mathbf{w}_i, \phi(\mathbf{x}) \rangle, \quad (20)$$

where

$$\mathbf{w}_i = \sum_{s \in SV_i} \alpha_s y_s (\phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)})) \quad (21)$$

Now, given that the definition of similarity uses scalar products instead of coordinates of weighting vectors, we can easily rewrite (18) in terms of the kernels used in the derivations explained in the previous subsection. The essential equality is:

$$\begin{aligned} \langle \mathbf{w}_i, \mathbf{w}_j \rangle &= \sum_{s \in SV_i} \sum_{l \in SV_j} \alpha_s \alpha_l y_s y_l \langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}_l^{(1)}) - \phi(\mathbf{x}_l^{(2)}) \rangle \\ &= \sum_{s \in SV_i} \sum_{l \in SV_j} \alpha_s \alpha_l y_s y_l \mathbf{K}(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)}, \mathbf{x}_l^{(1)}, \mathbf{x}_l^{(2)}) \end{aligned} \quad (22)$$

### 4.3 The clustering method

Once we have defined a reasonable similarity measure for preference criteria, we proceed to look for clusters of consumers with homogeneous tastes. In principle, we could use any available clustering algorithm. However, we avoided those methods, like k-means, that require frequent recomputations of the centroids of each cluster. The reason is that the updating of (22) would result computationally expensive. Additionally, we need a mechanism able to estimate an appropriate number of clusters directly from the data, without any explicit manual intervention.

Hence, we applied a *nonparametric pairwise* algorithm (Dubnov et al. 2002), although this is not the only possibility. The following paragraphs sketch a description of this algorithm as we used it in the experimental results reported in the last section.

Let  $\mathbf{M} = (M_{ij})$  be a square matrix where  $M_{ij}$  stands for the similarity between data points  $i$  and  $j$ ; in our case, data points are the vectorial representation of the preference criteria of consumers, and similarities are given by equation (18). In the following,  $\mathbf{M}$  will be called the *proximity matrix*. Then, matrix  $\mathbf{M}$  is transformed iteratively, following a two step procedure that converges to a two values matrix (1 and 0), yielding a bipartition of the data set into two clusters. Then, recursively, the partition mechanism is applied to each of the resulting clusters represented by their corresponding submatrices. To guarantee that

only meaningful splits take place, in (Dubnov et al. 2002) the authors provide a cross validation method that measures an index that can be read as a *significance level*; we will only accept splits whose level is above 0.90.

The basic iterative transformation uses the following formulae to go from iteration  $t$  to  $t+1$ :

$$\begin{aligned} V_{ij}(t+1) &= \frac{M_{ij}(t)}{\max\{M_{ik}(t): k\}} \\ M_{ij}(t+1) &= \frac{1}{2} \left( \sum_k V_{ik}(t+1) \log \frac{V_{jk}(t+1)}{\frac{1}{2}(V_{jk}(t+1) + V_{ik}(t+1))} + \sum_k V_{jk}(t+1) \log \frac{V_{ik}(t+1)}{\frac{1}{2}(V_{jk}(t+1) + V_{ik}(t+1))} \right) \end{aligned} \quad (23)$$

The first step normalizes the columns of the proximity matrix using the  $L_\infty$  norm; then the proximities are re-estimated using the Jensen-Shannon divergence. The idea is to formalize that two preference criteria are close (after these two steps) if they were both similar and dissimilar to analogous sets of criteria before the transformation.

## 5. Experimental results

To illustrate the performance of our proposals, we used two datasets. The first one is EachMovie, a dataset widely used in the field of collaborative filtering. The second dataset deals with ratings of beef meat consumers, and was previously used in (Díez et al. 2005). This is a dataset that was built in order to study the beef market preferences.

### 5.1 Clustering spectators according to their tastes about movies

The database EachMovie contains ratings of 1,628 movies provided by 72,916 people. The scale used has 6 values  $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$ , but less than only 2.4% of the possible values are filled. Thus, for instance, 11,651 people have not given any rating to any movie. To take into account explicitly the absence of ratings, that is so frequent in this collection, and as usual when dealing with EachMovie, we moved the scale of ratings to  $\{-1.5, -1, -0.5, 0.5, 1, 1.5\}$ , assuming zero as missing value.

As many authors do, in order to avoid the extreme sparsity of data, we considered a subset of EachMovie. We only considered movies with at least 1,000 ratings; there are 504 movies in such conditions. We likewise only took into consideration people who had submitted at least 275 ratings for these movies; this makes a total of 189 people.

To accomplish our setting for clustering, we considered 100 people as our sample to study possible similarity of preference criteria; i.e. we set  $m = 100$  (see equation (1)) and the remaining available ratings were used to define the input space  $X$  of the objects (movies) considered in our experiment. In other words, we represented each movie by a vector of 89 ( $= 189 - 100$ ) ratings. The distributions of ratings in the description of the movies, that is in the attribute space, and those provided by our 100 spectators are quite similar, see Table 2.

The main novelty in this setting with respect to typical collaborative filtering applications is that we use some consumers to describe the products. Notice that we only use “expert” consumers, those who evaluated a lot of movies, so in fact they form a kind of expert panel.

For each spectator  $p_i$  from the set of 100 people considered, we have a rating vector  $(r_i(\mathbf{x}) : \mathbf{x} \in S_i)$ , and we have to assemble a set of preference judgments  $PJ_i$  from it. Notice that  $|PJ_i|$  is of the order of  $O(|S_i|^2)$ . Here  $|S_i| \leq 504$ , so to avoid too big  $PJ_i$  datasets, we proceed as follows. For each movie  $\mathbf{x}^{(1)} \in S_i$  ranked by  $p_i$ , 10 more movies ranked by  $p_i$  with different ratings were randomly selected; then, if  $\mathbf{x}^{(2)}$  is one of such movies, we include  $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$  in  $PJ_i$  if and only if  $r_i(\mathbf{x}^{(1)}) > r_i(\mathbf{x}^{(2)})$ , otherwise, we include  $(\mathbf{x}^{(2)}, \mathbf{x}^{(1)})$  in  $PJ_i$ . Notice that in EachMovie there is no indication about rating sessions, therefore we assumed only one session, and hence all ratings are comparable.

Finally, we used two different kernels to appreciate their influence in the clusters obtained applying the method described in this paper. In Figure 2 we depict the tree of clusters obtained with linear kernel and polynomial kernel (degree=2). The overlapping between the clusters obtained by these two kernels is reported in Table 3.

The set of all preference judgments,

$$PJ = \cup(PJ_i : i = 1, \dots, m)$$

has a lot of *disagreements*. If, for each pair of movies, we choose the most frequent relative ordering in PJ, then the percentage of pairs disagreeing with that majority opinion is 25.92% (see Table 4). Notice that this percentage is the minimum training error for any algorithm trying to learn the preferences of the whole group of 100 spectators.

The clustering algorithm, with both kernels, has discovered groups of almost 50 people, whose disagreements are lower than the global percentage of disagreements: 17.4% for the linear kernel (li\_r), and 16.71% for the polynomial one (po\_r). These clusters are quite the same; since 42 spectators are present in these majority groups for both kernels (see Table 3). However, the other two smaller groups have internal disagreements similar to that of the whole group in both kernels. The reason for this behavior may be that these groups are made up of very peculiar spectators whose preference functions are difficult to learn.

The preference functions learned from the union of preference judgments of the members of each cluster have an estimation of misclassifications similar to the percentage of disagreements, especially in the case of the polynomial kernel, see Table 4.

## **5.2 Market segments in beef meat**

To illustrate our method with a real world application, we used a database described in (Sañudo et al. 2004; Díez et al. 2005). The data collects the sensory ratings of a panel of beef meat consumers about two aspects: tenderness, and acceptability. First we will describe the dataset, to conclude with a discussion of the implications of the results obtained by our method.

### **5.2.1 Beef meat dataset**

For this experience, 101 animals of 7 Spanish breeds were slaughtered to obtain two kinds of carcasses: lights, from animals with a live weight around 300–350 kg (light); and heavies, from animals at 530–560 kg. The set of animals was uniformly distributed by breeds and weights. Additionally, to test the influence of aging in consumers' appreciation, each piece of meat was prepared with 3 aging periods, 1, 7, and 21

days. Notice that these settings give rise to 303 (101 animal with 3 different aging periods) samples of meat.

On the other hand, the 7 breeds used constitute a wide representation of beef cattle. These breeds can be divided into four types: double muscled (DM, one breed), fast growth (FG, two breeds), dual purpose (DP, one breed), and unimproved rustic type (UR, three breeds). In Table 5, we show the average percentages of fat, muscle and bone for each breed. Notice that it is important to distinguish three kinds of fat: inter-muscular, subcutaneous, and intramuscular, that is included in the percentage of muscle. In the experimental results reported below, we used the carcass composition of each animal; the breed was only used to provide an explanation of the tastes of clusters. Additionally, each kind of meat was also described by a panel of 11 trained experts who rate 12 traits of products such as fibrosis, flavor, odor, etc.; we considered the average rate of each trait. The characterization of meat samples was completed with 6 physical features describing its texture.

The panel was organized in a set of tasting sessions, where a group of potential consumers assessed 3 or 7 instances of the available beef kinds. Frequently, each consumer only participated in a small number (sometimes only one) of tasting sessions, usually in the same day.

To apply the clustering method described in this paper, we first selected those people involved in our consumers' panel whose ratings gave rise to more than 30 preference judgments; these yielded us to consider a set of 171 panelists ( $m=171$ ) that tested from 9 to 14 samples of meat. Notice that this is just a subset of all available panelists since samples rated with the same rate did not produce preference judgment pairs. This individual treatment contrasts with the global one that we reported in (Luaces et al. 2004; Del Coz et al. 2004; Díez et al. 2004). In these prior works, we were interested in modeling the general opinion of consumers; so for each session, to summarize the opinions of consumers, we computed the mean of the ratings obtained by each piece of meat. Here, we will also use the polynomial kernel of degree 2 reported in those papers, since the nature of the learning problems is essentially the same.

Since there are 303 samples of meat, then the opinions of our panelists can only be estimated inducing a preference or ranking function (see section 4.1) using PJ datasets obtained applying equation (3). Notice that only such functions can be used in order to compare the preferences of different consumers; in



general, two arbitrary consumers have not tested samples of the same animal prepared with the same aging. However, it is possible to compare the preference functions of any couple of consumers as vectors in a high dimension space following the kernel based method of section 4.2.

The clustering algorithm returns the trees depicted in Figure 3. Split nodes achieved a confidence level of 91% for tenderness dataset, and 97% for acceptance. The leaves of these trees reached lower confidence levels, and therefore their splits were rejected.

In the scores reported in Table 6, we observe that both in acceptance and tenderness there are percentages of disagreements over 20%, while the clusters found decrease this amount in 5 points. Using the set of all preference judgments of each cluster, the preference functions learned (with a polynomial kernel of degree 2) have an estimation of misclassification around 20%.

### **5.2.2 Implications of beef meat results**

In general, it is well known that meat qualities are mainly the result of a set of complex factors. In this study, we focus our attention on two of them: breed and aging period. Specifically, we are interested in knowing if there are different groups of people who prefer some breeds to others or who prefer long to short aging periods. In fact, these are the most relevant features in the acceptance and tenderness appreciation of beef meat by consumers according to (Sañudo et al. 2004; Luaces et al. 2004).

To gain insight into the meaning of the preference criteria of each cluster, we used the ranking or preference functions to order the samples of meat; then we assessed 10 points to those samples included in the first decile, 9 to the second decile, and so on. Graphical representations of the average points obtained by each breed and each aging period are shown in Figure 4 (acceptance) and Figure 5 (tenderness); notice that the average score of all samples is 5.5. The results are quite the same if we use quartiles instead of deciles or any other division of the relative rankings of each cluster.

In the acceptance dataset (Figure 4a), let us emphasize the opposite role played by Retinta and Asturiana breeds: they were first and last (or almost last) in each cluster alternatively. In (Luaces et al. 2004) we used Boolean attributes to include the breed in the description of each sample, and then Retinta and

Asturiana were found to be the most relevant Boolean features in order to explain consumer's acceptance of meat. Additionally, these two breeds have significant differences in carcass composition (see Table 5). Notice that Asturiana breed is the only double muscled breed of the sample, and then it has the lowest values in percentages of subcutaneous and inter-muscular fat, and bone; while Retinta is the unimproved rustic breed with the highest percentages of fat and bone. Therefore, there are some reasons so as to assign opposite ratings to samples of these two breeds, although, in general, the final acceptance scorings rely on a complex set of features.

On the other hand, analyzing aging periods in both clusters, (Figure 4b) we can see that people in the left cluster prefer a 7 days aging period meanwhile in the right cluster longer aging periods have better acceptance.

In the tenderness dataset (Figure 5a), meat from Pirenaica and Retinta breeds are the tenderest for people in left cluster, however they are ranked in low positions in right cluster. We can say exactly the opposite of meat from Asturiana and Parda breeds. Again, Asturiana and Retinta breeds play opposite roles in each cluster.

In this case, we cannot appreciate differences between clusters' preferences about aging periods (Figure 5b): in both clusters, all breeds obtain higher tenderness scores as aging periods increase. This is not a surprisingly result, it is well documented that long aging periods give rise to more tender meat (Sañudo et al. 2004).

## **6. Conclusions**

We have presented a new algorithm to build clusters of preference criteria. Starting from a collection of preference judgments of different people, our algorithm discovers groups or clusters of people with closely related tastes; that is to say, people whose preference judgments can be merged in order to learn more reliable ranking functions able to express the preferences of the people involved. These clusters can be also interpreted as market segments. In building these clusters, the key insight is that ranking

functions, learned from the set of preference judgment of each person, codify the rationale or criteria for his or her preferences.

We believe that the contributions of this paper fall mainly within the applications of preference learning. Thus, clustering is useful for strengthening applications such as collaborative filtering, information retrieval or adaptive assistants. We illustrated the usefulness of the method presenting the experimental results achieved with a collaborative filter dataset (EachMovie), and with a dataset that gathers ratings of beef meat consumers. In this case, the clusters represent market segments with differentiated requirements from this kind of food products. It is straightforward the extension of this case to other fields that handle sensory data provided by consumer panels.

## Acknowledgements

The authors would like to thank: the Company Compaq for providing us with the EachMovie database (McJones, 1997); Thorsten Joachims for his SVM<sup>light</sup> (Joachims, 1998), and the authors of Spider (Weston et al., 2006), a MatLab toolbox that includes kernel based algorithms; these systems were used in the experiments reported in this paper.

## References

- Bahamonde, A.; Bayón, G. F.; Díez, J.; Quevedo, J. R.; Luaces, O.; del Coz, J. J.; Alonso, J.; Goyache, F. (2004). Feature subset selection for learning preferences: a case study. Proceedings of the 21st International Conference on Machine Learning, ICML 2004, Banff, Canada, pp. 49-56.
- Basu, C., Hirsh, H., and Cohen, W.W. (1998). Recommendation as classification: Using social and content-based information in recommendation. In Proc. AAAI. pp. 714-720.
- Branting, K.L. and Broos, P.S. (1997). Automated acquisition of user preferences. International Journal of Human-Computer Studies 46:55-77.

- Breese, J.S., Heckerman, D., and Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In Proc. Conference on Uncertainty in Artificial Intelligence. pp. 43-52.
- Buck, D., Wakeling, I., Greenhoff, K., and Hasted, A. (2001). Predicting paired preferences from sensory data. Food quality and preference 12:481-487.
- Cheung, W.K., Kwok, J.T., Law, M.H., and Tsui, K.C. (2000). Mining customer preference ratings for product recommendations using the support vector machine and the latent class model. In Proc. International Conference on Data Mining Methods and Databases for Engineering. pp. 601-610.
- Cohen, W.W., Shapire, R.E., and Singer, Y. (1999). Learning to order things. Journal of Artificial Intelligence Research 10:243-270.
- Corney, D. (2002). Designing food with bayesian belief networks. In Proc. International Conference on Adaptive Computing in Engineering Design and Manufacture. pp. 83-94.
- Crammer, K. and Singer, Y. (2001). Pranking with ranking. In Proc. Conference on Neural Information Processing Systems. pp. 641-647.
- Del Coz, J. J.; Bayón, G. F.; Díez, J.; Luaces, O.; Bahamonde, A.; Sañudo, C. (2004). Trait selection for assessing beef meat quality using non-linear SVM. Proceedings of the 18th Annual Conference on Neural Information Processing Systems (NIPS 2004). Vancouver, British Columbia, Canada, pp. 321-328.
- Díez, J.; Bayón, G.F.; Quevedo, J. R.; del Coz, J. J.; Luaces, O.; Alonso, J.; Bahamonde, A. (2004). Discovering relevancies in very difficult regression problems: applications to sensory data analysis. In Proceedings of the European Conference on Artificial Intelligence (ECAI '04), Valencia, Spain, pp. 993-994.
- Díez, J.; del Coz, J. J.; Sañudo, C.; Albertí, P.; Bahamonde, A. (2005). A Kernel Based Method for Discovering Market Segments in Beef Meat. Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, ECML/PKDD'2005, Porto, Portugal, pp. 462-469.
- Dubnov, S.; El-Yaniv, R.; Gdalyahu, Y.; Schneidman, E.; Tishby, N.; Yona, G.. A New Nonparametric Pairwise Clustering Algorithm Based on Iterative Estimation of Distance Profiles. Machine Learning, 47 (2002) 35-61
- Dumais, S., Bharat, K., Joachims, T., and Weigend, A. (2003). Workshop on implicit measures of user interests and preferences. In ACM SIGIR Conference, Toronto, Canada.
- Fiechter, C.N. and Rogers, S. (2000). Learning subjective functions with large margins. In Proc. International Conference on Machine Learning. pp. 287-294.

- Freund, Y., Iyer, R., Schapire, R.E., and Singer, Y. (1998). An efficient boosting algorithm for combining preferences. In Proc. International Conference on Machine Learning. pp. 170-178.
- Goldberg, D., Nichols, D., Oki, B.M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 35 (12):61-70.
- Goyache, F., Bahamonde, A., Alonso, J., López, S., del Coz J.J., Quevedo, J.R., Ranilla, J., Luaces, O., Alvarez, I., Royo, L., and Díez, J. (2001). The usefulness of artificial intelligence techniques to assess subjective quality of products in the food industry. *Trends in Food Science and Technology* 12 (10):370-381.
- Herbrich, R.; Graepel, T.; and Obermayer, K.: Large margin rank boundaries for ordinal regression. In A. Smola, P. Bartlett, B. Scholkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, 115–132. MIT Press, Cambridge, MA, (2000)
- Hofmann, T and Puzicha, J. (1999) Latent class models for collaborative filtering. In Proc. International Joint Conference on Artificial Intelligence. pp.688-693.
- Joachims, T. (1998). Making large-scale SVM learning practical. *Advances in Kernel Methods - Support Vector Learning*, MIT Press, Cambridge MA. Chapter 11.
- Joachims, T. (2002). Optimizing search engines using clickthrough data. In Proc. ACM Conference on Knowledge Discovery and Data Mining.
- Luaces, O.; Bayón, G.F.; Quevedo, J.R.; Díez, J.; del Coz, J.J.; Bahamonde, A. (2004). Analyzing sensory data using non-linear preference learning with feature subset selection. *Proceedings of the 15th European Conference of Machine Learning, (ECML 04)*. Pisa, Italia, pp. 286-297.
- McJones, P. (1997). EachMovie collaborative filtering dataset. DEC (now Compaq) Systems Research Center. <http://www.research.compaq.com/SRC/eachmovie/>.
- Murray, J.M., Delahunty, C.M., and Baxter, I.A. (2001). Descriptive sensory analysis: Past, present and future. *Food Research International* 34:461-471.
- Pazzani, M.J. (1999). A framework for collaborative filtering, content-based and demographic filtering. *Artificial Intelligence Review* 13 (5-6):393-408.
- Resnick, P. and Varian, H. (1997). Introduction to special section on recommender systems. *Communications of the ACM* 40(3):56-58.

- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of networks. In Proc. Conference on Computer Supported Cooperative Work. pp. 175-186.
- Sañudo, C.; Macie, E.S.; Olleta, J.L.; Villarroel, M.; Panea, B.; Albertí, P.. The effects of slaughter weight, breed type and ageing time on beef meat quality using two different texture devices. *Meat Science*, 66 (2004), 925–932
- Shardanand, U. and Maes, P. (1995). Social information filtering: Algorithms for automatic “Word of Mouth”. In Proc. Conference on Human Factors in Computing Systems. pp. 210-217.
- Shashua, A. and Levin, A. (2002). Ranking with large margin principle: Two approaches. In Proc. Neural Information and Processing Systems.
- Tesauro, G. (1989). Connectionist learning of expert preferences by comparison training. In Proc. Neural Information and Processing Systems. pp. 99-106.
- Ungar, L.H. and Foster, D.P. (1998). Clustering methods for collaborative filtering. In Proc. AAAI’98 Workshop on Recommendation Systems. pp. 112-125.
- Utgoff, J. P. and Clouse, J. (1991). Two kinds of training information for evaluation function learning. In Proc. AAAI’91. pp. 596-600.
- Vapnik, V. (1998). *Statistical learning theory*. John Wiley, New York.
- Weston, J.; Elisseeff, A.; BakIr, G.; Sinz, F. (2006) SPIDER: object-orientated machine learning library. <http://www.kyb.tuebingen.mpg.de/bs/people/spider/>

**Table 1.** Sensory data collected from panels of experts and consumers. Each product is described by expert assessments ( $Att_i$ ) in addition to other ( $O-Att_i$ ) chemical, biological or physical analysis outputs

Expert sensory appreciations				Other descriptive attributes		Consumer preferences			
Expert-1		Expert-k							
$Att_1$	...	$Att_m$	...	$Att_1$	...	$Att_m$	O- $Att_1$	...	O- $Att_n$
<num>	...	<num>	...	<num>	...	<num>	<i>	<Id>	<num>
...	...	...	...	...	...	...	...	...	...
<num>	...	<num>	...	<num>	...	<num>	<i>	<Id>	<num>

**Table 2.** Rating distribution in data sets used for the experiments

Movies' Rating	In attribute space		In the 100 spectators' ratings	
	Number	Percentage	Number	Percentage
0	4,392	17.02%	6,665	19.32%
0.2	1,955	7.57%	2,651	7.68%
0.4	3,813	14.77%	4,801	13.92%
0.6	6,373	24.69%	8,193	23.75%
0.8	5,774	22.37%	7,471	21.66%
1	3,503	15.57%	4,716	13.67%
Totals	25,810		34,497	

**Table 3.** Number of common spectators in clusters obtained with the linear kernel and the polynomial kernel of degree 2

		Polynomial of degree 2			
		po_l_l	po_l_r	po_r	
Lineal	li_l_l	12	12	3	27
	li_l_r	5	17	2	24
	li_r	7	0	42	49
		24	29	47	

**Table 4.** Number of disagreements (minimum possible training error) and classification error estimations in each cluster depending of the kernel used

Kernel	disagreements	cluster	classification	
			disagreements	errors
linear	25.92%	li_l_l	27.35%	32.71%
		li_l_r	25.78%	30.62%
		li_r	17.40%	19.18%
polynomial	25.92%	po_l_l	23.8%	24.68%
		po_l_r	30.36%	30.86%
		po_r	16.71%	17.00%

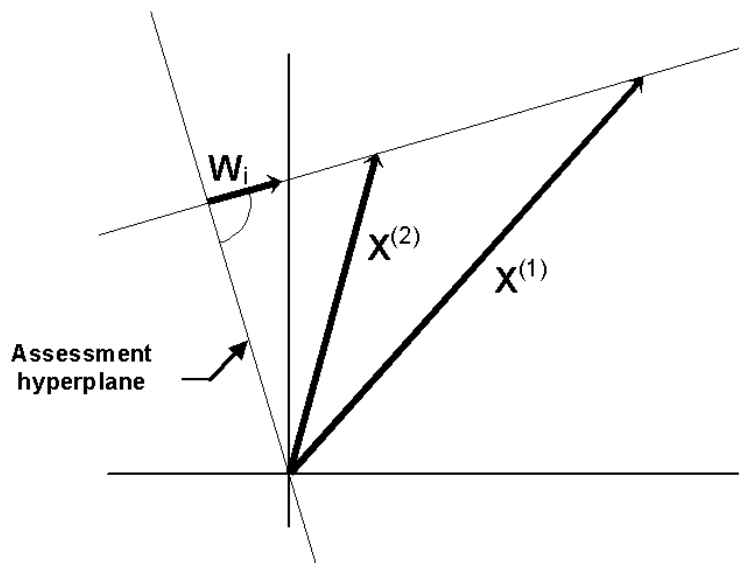
**Table 5.** Carcass compositions of 7 Spanish beef breeds used in the experiment

Breed		Fat %		Bone	Muscle	Intramuscular
Name	Type	inter-muscular	subcutaneous	%	%	fat %
Asturiana Valles	DM	4.77	0.89	16.00	78.34	0.90
Avileña	UR	13.17	3.53	19.25	64.05	2.28
Morucha	UR	12.46	3.46	19.28	64.80	2.10
Parda Alpina	DP	9.65	2.32	20.86	67.17	1.82
Pirenaica	FG	9.02	3.01	17.33	70.63	1.48
Retinta	UR	14.16	4.75	20.89	60.20	2.13
Rubia Gallega	FG	5.73	1.20	16.56	76.52	1.12

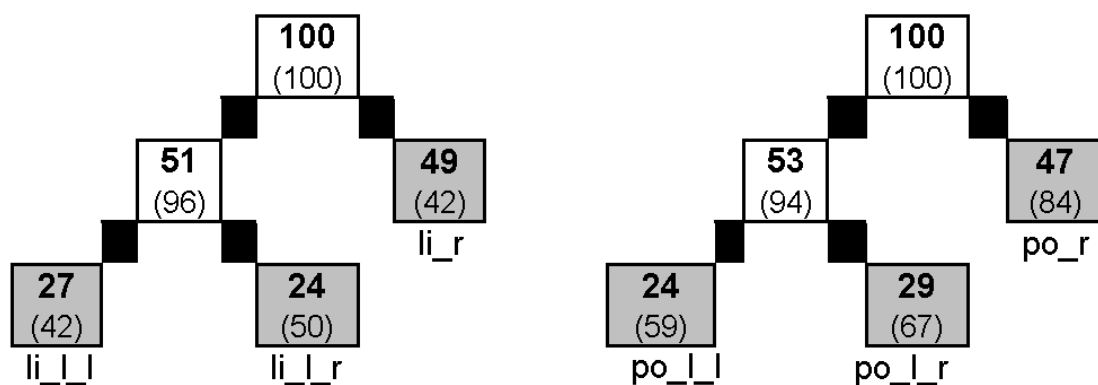
**Table 6.** Number of disagreements (minimum possible training error) and classification error estimations in each cluster for acceptance and tenderness

Dataset	disagreements	cluster	PJ	classification	
				disagreements	errors
acceptance	21.41%	left	1927	16.19%	19.20%
		right	2150	17.07%	21.12%
tenderness	20.25%	left	2487	15.96%	19.38%
		right	2432	15.21%	19.59%





**Fig. 1.** The assessment of each object is proportional to the distance of the vector that represents it to the (assessment) hyperplane perpendicular to  $w_i$ . In the picture,  $x^{(1)}$  is preferable to  $x^{(2)}$ , since it is further



**Fig. 2.** Trace of the clustering algorithm in the EachMovie dataset. The left hand side corresponds to the linear kernel while the right side represents the result obtained with the polynomial kernel of degree 2. In each node we report the number of spectators in **bold**, and the confidence level required to split the node (in parenthesis). The leaves are labelled for further reference

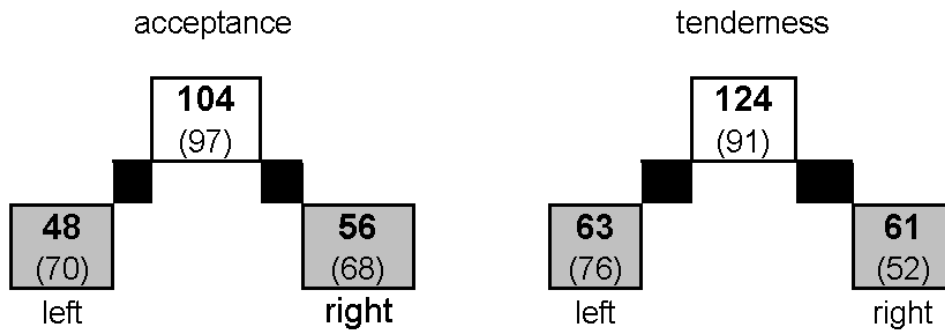


Fig. 3. Trace of the clustering algorithm in the beef meat dataset

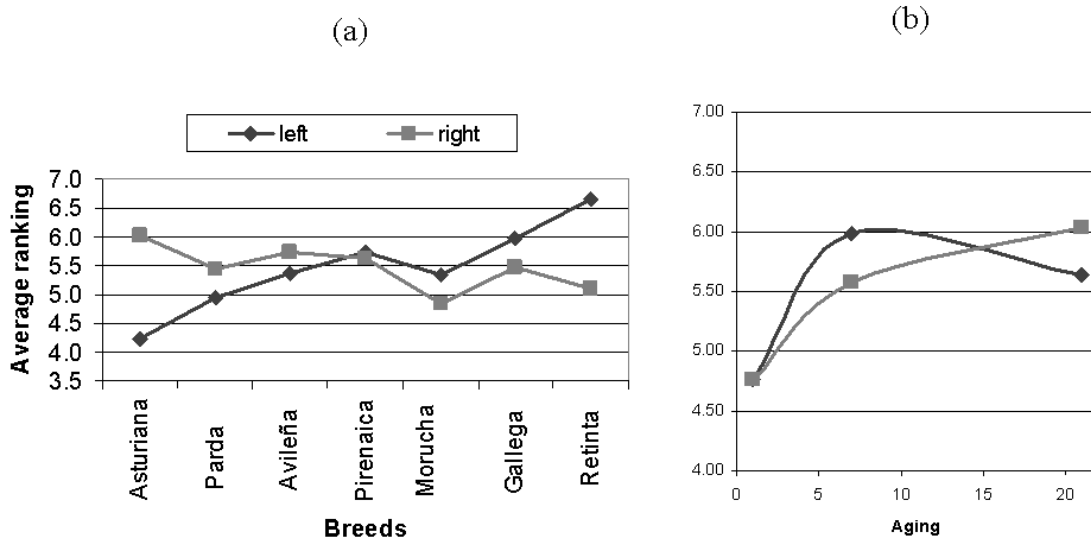


Fig. 4. Acceptance of beef meat. Average ranking scores for each breed and aging period

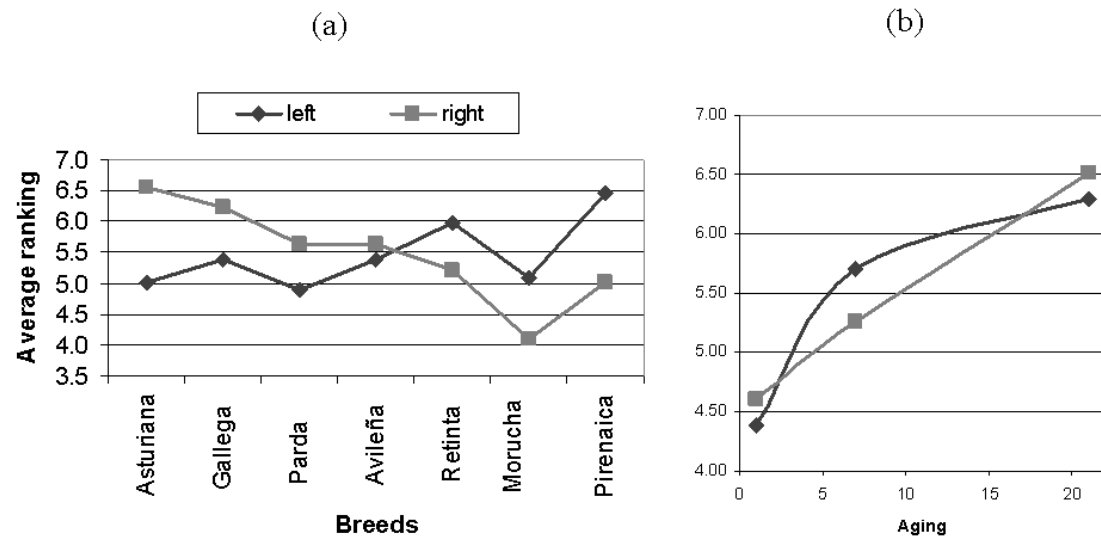


Fig. 5. Tenderness of beef meat. Average ranking scores for each breed and aging period