

Ordinal classification of depression spatial hot-spots of prevalence

M. Pérez-Ortiz[†], P.A. Gutiérrez[†], C. García-Alonso[‡], L. Salvador-Carulla*, J.A. Salinas-Pérez[‡], C. Hervás-Martínez[†]

[†] *Department of Computer Science and Numerical Analysis
University of Córdoba, Córdoba, Spain
{i82perom, pagutierrez, chervas}@uco.es*

[‡] *Department of Management and Quantitative Methods
ETEA, Business Administration Faculty. University of Córdoba, Córdoba, Spain
{cgarcia, jsalinas}@etea.com*

* *PSICOST Scientific Association
Jérez de la Frontera, Cádiz, Spain
luis.salvador@telefonica.net*

Abstract—In this paper we apply and test a recent ordinal algorithm for classification (Kernel Discriminant Learning Ordinal Regression, KDLOR), in order to recognize a group of geographically close spatial units with a similar prevalence pattern significantly high (or low), which are called hot-spots (or cold-spots). Different spatial analysis techniques have been used for studying geographical distribution of a specific illness in mental health-care because it could be useful to organize the spatial distribution of health-care services. Ordinal classification is used in this problem because the classes are: spatial unit with depression, spatial unit which could present depression and spatial unit where there is not depression. It is shown that the proposed method is capable of preserving the rank of data classes in a projected data space for this database. In comparison to other standard methods like C4.5, SVMRank, Adaboost, and MLP nominal classifiers, the proposed KDLOR algorithm is shown to be competitive.

Keywords—ordinal classification; ordinal regression; spatial distribution of illnesses; geographical information systems; kernel discriminant learning.

I. INTRODUCTION

The study of the spatial-geographical distribution of relevant phenomena is especially important in health care research and in health economics. This scientific interest is mainly justified by the potential existence of spatial dependence of the prevalence in specific illnesses. The use of different spatial analysis techniques for studying the geographical distribution of the prevalence of specific illnesses in mental health-care has shown many and severe discrepancies [1]. The resulting disparity makes the validity of the results obtained difficult in decision-making, but this drawback can be artificially masked if only one method is used.

The relationships between illness and socio-economics factors have been studied by many epidemiological studies. For example, the relationship between cancer registrations,

hospital discharge rate and socio-economics and deprivation index was studied using weighted linear regression analysis (weighted by population size) and Cochran-Armitage trend test [2]; or logistic regression was used to link socio-demographic variables and incidence, mortality and case survival trends for cancer of unknown primary site [3]. Neural networks have been used to study the connections between risk factors (including socio-demographic) and illnesses as the West Nile Virus [4] or prostate cancer [5].

Econometric spatial analysis includes different methods like classical poisson analysis [6], local indicators of spatial aggregation [7], [8], artificial intelligence based ones [9] and others.

In order to deal with the disparities observed when different spatial methods are used and compared [1], a very recent ordinal classification algorithm [10] has been implemented in this paper to identify geographic areas that group close spatial units, where the prevalence of selected illnesses (or their autocorrelation scores) was highly and significantly concentrated: hot-spots. These areas can be then located in a map using standard tools in Geographical Information Systems. Techniques for spatial analysis of geographical areas with a statistically significant high or low autocorrelated prevalence of any given disorder are especially interesting in health resources planning and management because they can be used to both highlight relevant concentrations of illnesses as well as to efficiently locate health-care units. The relationship between the spatial allocation of hot-spots and the corresponding socio-economic characteristics of the spatial units within them is a serious and arguable problem. The objective of this paper is to find the relationships between socio-economic data that could influence the existence and location of hot-spots. The geographical distribution —years 2004 and 2007— of the depression prevalence in Andalusia (south of Spain) is used to demonstrate the validity and

usability of our methodological approach.

Ordinal regression has a wide range of applications in areas where the human evaluation plays an important role, for example: psychology, medicine, information retrieval, etc., and, similarly, this method can be applied to other topics. In ordinal classification, the variable to predict is not numeric or nominal, but ordinal, so the categories have a natural order. The major problem with this type of classification is that there is not a precise notion of the distance between classes. For example, on a scale of ranges we know that 3 is closer to 2 than 5, but on an ordinal scale as it could be {very good, good, medium, bad, very bad}, how do we know that medium is closer to good than very bad? Traditionally, ordinal classification problems have been solved by a multinomial approach, i.e. ignoring the order among categories. But this is not a good simplification, since when classifying, the error obtained by confusing the very bad class with the medium one is equal to that of confusing the very bad class with the very good one (when in an ordinal scale, the latter should be more penalized).

Most current methods in ordinal regression are based in Support Vector Machines (SVMs) [11] but these methods ignores the global information of data and have a high computational complexity. In contrast, linear discriminant methods (Linear Discriminant Analysis, LDA) and its kernel version (Kernel Discriminant Analysis, KDA) covers these deficiencies [12] but their use for this purpose is very recent. Because of this reason, KDA method adapted for ordinal regression has been applied in this paper to the problem of spatial distribution of illnesses.

This paper is organized as follows: a brief analysis of the basis of the presented algorithm is given in Section II. The spatial hot-spots database and its features are presented in Section III. Finally, we present the results comparing the algorithm with C4.5, SVMRank, Adaboost, and MLP algorithms in Section IV and Section V summarises the conclusions of our work.

II. METHOD

The presented algorithm (Kernel Discriminant Learning Ordinal Regression) [10] is an extension of the well-known Kernel Discriminant Analysis (KDA), but applied to ordinal regression. The main goal can be described as finding the optimal linear projection for classification (from which different classes can be well separated), preserving at the same time the ordinal information of classes, i.e. the average projection of the samples from the higher rank classes should be larger than that of lower rank classes.

Therefore, this algorithm analyses three objectives: maximizing the between-class distance, minimizing the within-class distance and ensuring the ordinal information of the different classes. Given a set of training samples $(\mathbf{x}_i, y_i) \in R^l \times R$, $i = \{1, \dots, N\}$, where $\mathbf{x}_i \in R^l$ denotes inputs and

$y_i \in \{1, 2, \dots, K\}$ denotes the corresponding ordinal class labels, K is the defined number of classes, N is the sample size and N_k is the number of samples for the k -th class, a between-class scatter matrix (\mathbf{S}_b) and a within-class scatter matrix (\mathbf{S}_w) is defined as follows:

$$\mathbf{S}_b = \frac{1}{N} \sum_{k=1}^K N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^T, \quad (1)$$

$$\mathbf{S}_w = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{N_k} (\mathbf{x}_i - \mathbf{m}_k)(\mathbf{x}_i - \mathbf{m}_k)^T, \quad (2)$$

where $\mathbf{m}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{x}_i$ and $\mathbf{m} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$.

The objectives presented can be achieved by the maximization of the so called Rayleigh coefficient, a classical and well-know technique that solves this type of problem:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}, \quad (3)$$

where \mathbf{w} is the projection we are looking for.

A. Maximization of the Rayleigh coefficient

In order to maximize the previous Rayleigh coefficient, the method is reformulated. One should take into account that maximizing $\frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}$ is the same than minimizing $\frac{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}$. As $\mathbf{w}^T \mathbf{S}_b \mathbf{w}$ is an scalar, it is associated to a variable ρ , which represents the minimum difference of the projected means between consecutive classes, so if $\rho > 0$ the projected means are ranged according to the ordinal scale. Finally, the original optimization problem is transformed into the following one, which tries to minimize the variances of the data for the same classes while simultaneously extending the difference between the projected means of two neighbouring classes:

$$\min J(\mathbf{w}, \rho) = \mathbf{w}^T \mathbf{S}_w \mathbf{w} - C\rho, \quad (4)$$

with the restriction $\mathbf{w}^T (\mathbf{m}_{k+1} - \mathbf{m}_k) \geq \rho$.

C is a penalty coefficient which can be adjusted using a cross-validation method. If the value of C is set appropriately, the constraint $\rho > 0$ is automatically satisfied.

Taking into account that the original problem may be in a finite feature space, it could happen that the set to discriminate is not linearly separable. So the original finite feature space is usually transformed into a higher dimensional one, allowing then a simpler separation of the patterns (linear). Therefore, the kernel trick is used.

A kernel learning algorithm is used for solving non-linear regression problems and the procedure is to map the original training data via a mapping function ϕ into a higher dimensional feature space ($\phi : \mathbf{x} \rightarrow \phi(\mathbf{x})$), where each coordinate corresponds to one feature of the data items. Then, the data will be linearly separable and, as a result,

the idea of Linear Discriminant Analysis can be applied. So, the initial problem is reformulated as:

$$\mathbf{S}_w = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{N_k} (\phi(\mathbf{x}_i) - \mathbf{M}_k)(\phi(\mathbf{x}_i) - \mathbf{M}_k)^T,$$

$$\mathbf{S}_b = \frac{1}{N} \sum_{k=1}^K N_k (\mathbf{M}_k - \mathbf{M})(\mathbf{M}_k - \mathbf{M})^T,$$

where $\mathbf{M}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \phi(\mathbf{x}_i)$, and $\mathbf{M} = \frac{1}{N} \sum_{i=1}^N \phi(\mathbf{x}_i)$.

The problem is now solved by optimizing a $\boldsymbol{\beta}$ vector and defining the corresponding $\mathbf{w} = \sum_{i=1}^N \beta_i \phi(\mathbf{x}_i)$ vector which will determine the classification hyperplane. Consequently, the optimization problem can be turned into the following one:

$$\begin{aligned} \min J(\boldsymbol{\beta}, \rho) &= \boldsymbol{\beta}^T \mathbf{H} \boldsymbol{\beta} - C\rho, \\ \text{s.t. } \boldsymbol{\beta}^T (\mathbf{M}_{k+1} - \mathbf{M}_k) &\geq \rho, \end{aligned} \quad (5)$$

where $\mathbf{H} = \sum_{k=1}^K \mathbf{P}_k (\mathbf{I} - \mathbf{1}_{N_k}) \mathbf{P}_k^T$, \mathbf{P}_k is an $N \times N_k$ matrix with $(\mathbf{P}_k)_{i,j} = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle$, and $\mathbf{x}_j \in X_k$.

To solve it, Lagrange multipliers can be applied, a method for optimizing functions of several variables subject to constraints. The initial function and the constraints are joint in the following unique function:

$$L(\boldsymbol{\beta}, \rho, \boldsymbol{\alpha}) = \boldsymbol{\beta}^T \mathbf{H} \boldsymbol{\beta} - C\rho - \sum_{k=1}^{K-1} \alpha_k \left\{ \boldsymbol{\beta}^T (\mathbf{M}_{k+1} - \mathbf{M}_k) - \rho \right\}.$$

The α_k coefficients are the Lagrange multipliers ($\alpha_k \geq 0$). To find the minimum or maximum of this function, L must be derived with respect to $\boldsymbol{\beta}$ and ρ :

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0 \implies \boldsymbol{\beta} = \frac{1}{2} \mathbf{H}^{-1} \sum_{k=1}^{K-1} \alpha_k (\mathbf{M}_{k+1} - \mathbf{M}_k), \quad (6)$$

$$\frac{\partial L}{\partial \rho} = 0 \implies \sum_{k=1}^{K-1} \alpha_k = C. \quad (7)$$

So, by joining (6) and (7), the following function should be optimized:

$$\begin{aligned} \min f(\boldsymbol{\alpha}) &= \\ &= \sum_{k=1}^{K-1} \alpha_k (\mathbf{m}_{k+1} - \mathbf{m}_k)^T \mathbf{S}_w^{-1} \sum_{k=1}^{K-1} \alpha_k (\mathbf{m}_{k+1} - \mathbf{m}_k) \\ \text{s.t. } \alpha_k &\geq 0, k = 1, \dots, K-1, \text{ and } \sum_{k=1}^{K-1} \alpha_k = C. \end{aligned} \quad (8)$$

Algorithm KDLOR

- **Input:** Training patterns (\mathbf{Tr}), Training targets (\mathbf{Trg}), Test patterns (\mathbf{Ts}).
- **Output:** Predicted test targets (\mathbf{Tsg}).
 - 1) Apply Gaussian kernel to training patterns \mathbf{Tr} (\mathbf{K}_{Tr}) and test patterns \mathbf{Ts} (\mathbf{K}_{Ts}).
 - 2) Calculate \mathbf{H} and \mathbf{Q} matrices.
 - 3) Solve the optimization problem with linear constraints.
 - 4) Calculate the optimal projection $\boldsymbol{\beta}$ and the thresholds b_k for discriminating the classes.
 - 5) Finally, calculate the class membership (\mathbf{Tsg}) using the projection, the thresholds and the kernel matrix for the test patterns.

Figure 1. Different steps of the KDLOR algorithm.

B. Optimization problem

The optimization of the function is not specified in the work where KDLOR is proposed [10]. This optimization problem is a convex Quadratic Programming (QP) with linear constraints. For the optimization of the function, we reformulate it in the following canonical form of the QP problems:

$$\min F(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x},$$

with the constraints $A\mathbf{x} \leq b$ and $E\mathbf{x} = d$.

Problem of equation (8) can be solved by using the following \mathbf{Q} matrix:

$$Q_{ij} = 2(\mathbf{m}_{i+1} - \mathbf{m}_i)^T \mathbf{S}_w^{-1} (\mathbf{m}_{j+1} - \mathbf{m}_j).$$

It is not necessary to use the vector \mathbf{c} , so we can fill it with zeros. The final function is:

$$f(x) = (\alpha_1 \quad \dots \quad \alpha_{K-1}) \mathbf{Q} \begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_{K-1} \end{pmatrix}. \quad (9)$$

The optimization problem has been solved with the Quadratic Programming function implemented in Matlab (named 'quadprog'). For more information see [13].

The kernel selected for this purpose is the Gaussian one, $K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}\right)$ where σ^2 is the variance (width of the kernel). If the problem is ill-posed, \mathbf{S}_w becomes singular. To solve this, a regularization method can be used (adding a constraint $u > 0$ to the diagonal elements of \mathbf{S}_w). The value of u is fixed by a cross-validation method too.

After obtaining the optimal projection $\boldsymbol{\beta}$ by substituting α_k into (6), the rank of an unseen input vector \mathbf{x} can be predicted by the following decision rule:

$$f(\mathbf{x}) = \begin{cases} K, & \text{if } \mathbf{w} \cdot \mathbf{x} - b_{K-1} > 0, \\ \max_k \{\mathbf{w} \cdot \mathbf{x} - b_k < 0\}, & \text{otherwise.} \\ k \in \{1, \dots, K-1\} \end{cases} \quad (10)$$

where $b_k = \frac{\mathbf{w}(\mathbf{m}_{i+1} + \mathbf{m}_i)}{2}$ with $k = 1, \dots, K-1$.

Taking into account all these steps, the KDLOR algorithm is explained in Figure 1, where \mathbf{K}_{Tr} and \mathbf{K}_{Ts} are the training and test kernel matrices.

III. SPATIAL HOT-SPOTS

Each Spatial Unit (SU) is classified in one of the three classes: out of the hot-spots (class #1), on the border (class #2) or in the hot-spots (class #3), resulting in a very unbalanced problem. This fact is due to the scarce number of municipalities in classes #2 and #3 in comparison to the corresponding for class #1. In big regions, where the latter can be 50 times greater than the former ones, classification methods can not explain the classes. In order to remove those SUs in class #1 very far away from the identified hot-spots, a statistical procedure has been designed for analysing only the ones that can be considered relatively close from a geographical point of view.

When autocorrelation scores were calculated, the geographical distance between two SUs was normalised within an integer [1, 10] range, with a 1 value for those ones that are neighbours and 10 for those that are very far away. Taking into account the results of this procedure, we then try to eliminate those SUs which are isolated. In this way, those SUs which distances to all the SUs in classes #2 and #3 (on the border and in the hot-spots) are greater than a specific threshold value d_i for a percentage of patterns higher than pd_i have been removed from the database. This procedure guaranteed that only comparable SUs were included in the classification problem. Once a SU is selected, its characteristics were used to explain the assigned class.

According to this methodology in 2004 and 2007, our procedure classified the SUs—municipalities—in: out of the hot-spots (class #1), on the border of the hot-spots (class #2) and in the hot-spots (class #3). In Andalusia, there were 770 SUs that can be described socio-economically, but only a few can be included in depression hot-spots (9 and 7, respectively, in the 2004 and 2007 databases, Table 1) or on their borders (13 and 23 respectively in the 2004 and 2007 databases, Table 1). When the ownership of a specific depression-based class should be explained by socio-economic variables, the structure of the databases makes the classification problem infeasible. In order to remove very isolated SUs, d_i was established equal to 5 and pd_i was fixed to 70%. The resulting geographical areas are shown in Figure 2).

As expected from an epidemiological point of view, the location of the three classes is relatively constant, with some relevant differences, throughout the time span. Their socio-economic characteristics in 2004 and 2007 are summarised in Table I. Taking into account that there are non significant differences between their means, the resulting classification model can be considered specially complex.

The selected socio-economic characteristics are: Unemployment Rate (UR, persons/100 active persons), Illiterate Persons Rate (ILR, persons older than 10 years of age/100 persons older than 10), Specialised Workers Rate (SWR, number of specialized workers/100 workers), Households

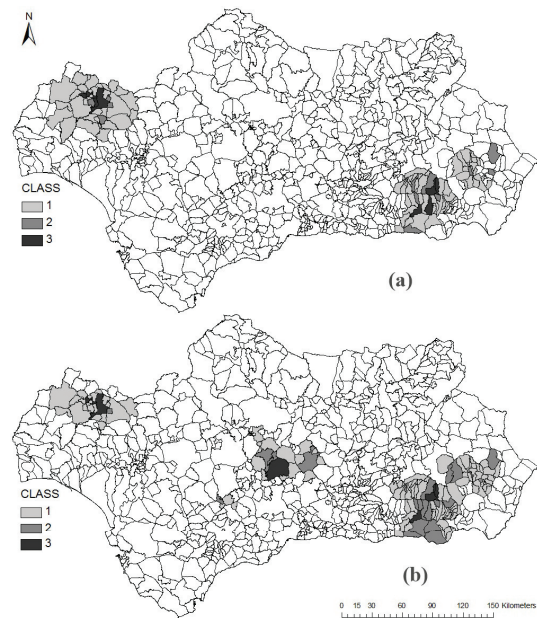


Figure 2. Spatial distribution of (a) 2004 and (b) 2007 years.

without Car Rate (HCR, number of households without cars/100 households), Inactive Persons Rate (IPR, Number of inactive persons—younger than 14 and older than 65 years of age—/100 potentially active persons), New Foreigners Rate (NFR, number of foreigners—14 to 49 years old—that lived outside Spain in 1991/100 foreigner persons) and, finally, Deprivation Index (DI). The hypothesis of this research is to demonstrate the existence of a relationship between them and the geographical location of depression hot-spots.

IV. RESULTS

Three evaluation metrics are considered to quantify the accuracy of N predicted ordinal scales: Correct Classification Ratio (CCR), Mean absolute error (MAE) and Minimum sensitivity (MS).

The spatial structure of the year 2004 was considered as the training set and the corresponding one for 2007 was used as the test set, for all the classification methods compared, because of the temporary linearity.

We have compared the algorithms with other four state-of-the-art methods well known in the literature. Some of them have been configured and run in WEKA [14]. The methods used for comparison are:

- C4.5 is the standard algorithm for training classifiers based on a pruned or unpruned decision tree. The confidence factor used for pruning is 0.1 and the minimum number of instances per leaf is 2.
- SVM-Rank applies the Extended Binary Classification (EBC) method to SVM.

Table I
BASIC STATISTICS OF THE INPUT VARIABLES

	UR(1)	ILR(2)	SWR(3)	HCR(4)	IPR(5)	NFR(6)	DI(7)
Year: 2004; Class #1: Out of the hot-spot (79 municipalities)							
Minimum	2.10	0.00	24.00	22.30	78.50	0.00	-0.33
Maximum	57.70	15.60	76.00	66.70	322.00	4.10	4.02
Average	25.33	4.88	50.39	45.97	167.70	0.35	1.32
STD(8)	12.00	2.92	8.44	8.03	51.78	0.71	0.77
Year: 2004; Class #2: On the border of the hot-spot (13 municipalities)							
Minimum	8.30	2.60	36.00	14.70	71.00	0.00	0.23
Maximum	34.20	10.20	71.10	60.80	215.00	3.60	2.26
Average	19.67	5.35	51.62	40.95	144.76	0.59	1.17
STD	7.67	2.62	9.96	11.89	52.73	1.10	0.70
Year: 2004; Class #3: Hot-spots (9 municipalities)							
Minimum	13.10	1.10	33.50	30.20	101.00	0.00	0.09
Maximum	37.80	8.40	64.60	46.30	218.00	0.60	2.34
Average	25.83	4.46	51.24	41.74	146.33	0.22	1.29
STD	8.34	2.90	9.30	4.81	43.68	0.22	0.68
Year: 2007; Class #1: Out of the hot-spot (102 municipalities)							
Minimum	2.10	0.00	24.00	21.30	65.80	0.00	-0.33
Maximum	69.60	12.70	81.40	66.70	381.00	16.10	4.02
Average	25.18	4.94	52.57	42.67	158.22	0.84	1.35
STD	13.76	2.60	9.17	9.21	57.24	2.14	0.78
Year: 2007; Class #2: On the border of the hot-spot (23 municipalities)							
Minimum	5.10	1.50	35.80	27.00	64.80	0.00	-0.07
Maximum	43.50	10.20	75.60	60.80	264.00	10.90	2.70
Average	21.42	5.92	55.10	39.42	143.12	1.01	1.41
STD	10.40	2.45	10.02	8.39	57.92	2.34	0.79
Year: 2007; Class #3: Hot-spots (7 municipalities)							
Minimum	14.80	1.10	33.50	25.80	91.70	0.00	0.09
Maximum	37.80	7.80	57.70	46.30	218.00	0.90	2.34
Average	26.36	4.53	49.36	39.04	147.39	0.34	1.25
STD	8.34	2.70	8.96	7.85	48.31	0.34	0.77

- AdaBoost100 is the application of AdaBoost.M1 algorithm with 100 maximum of iterations and a classifier tree as the base learner. AdaBoost.M1 gives higher weights to the misclassified instances in order to force the new classifiers to concentrate on those. The threshold for weight pruning is set to 100 and the number of iterations to 100.
- MultiLayer Perceptron (MLP) is a standard neural network classifier that uses back-propagation to calculate the weights. The nodes in this network are all sigmoid. The learning rate is set to 0.1 and the momentum applied to the weights is 0.2.

Extended details about all these algorithms can be found in [15]. Several parameters have been tested and chosen for each method, those which performs the best accuracy of the algorithms. Moreover, the linear version of KDLOR (without applying the kernel method) has been considered for comparison purposes, which will be called Linear Discriminant Analysis for Ordinal Regression (LDAOR).

The parameters of KDLOR algorithm has been estimated via a crossvalidation process, the selected value for the penalty coefficient is $C = 1$, the value for the width of the kernel is $\sigma = 1$ and the constant value for avoiding singularity $u = 0.01$. The algorithm was implemented and run using Matlab software.

The results of all the different methods are included in Table II. From these results, several conclusions can be derived. Firstly of all, the success rates, the MS and the MAE

are acceptable for some methods (KDLOR, SVM-Rank and AdaBoost100). However, there are very low performance methods (LDAOR for example), for which the classification task is not correctly done. The necessity of applying a kernel is demonstrated by these results, together with the high non-linearity of the considered dataset. It is also interesting to consider the importance of taking into account the order when evaluating these methods: for example, KDLOR is a 18.73% better than C4.5 when considering MZE but this difference is higher when considering MAE (25.65%). In this way, accurate classifiers can be masking sometimes a low MAE performance.

To evaluate the complete classification task, the confusion matrices have been included in Table III. In these matrices, it is clearly shown how the class #2 is easily confused with the adjacent classes (#1 and #3). This problem is due to the high overlapping of the class and the complex associated boundaries. The low number of patterns in some classes is also problematic (especially in class #3, but also in class #2). Some models tend to ignore these classes (e.g. MLP does not classify any pattern in class #3, and LDAOR does not classify any pattern in class #2). In general, the better balance of classification is achieved by using KDLOR.

V. CONCLUSIONS

This paper has introduced the application of a very recent ordinal classifier for predicting spatial hot-spots, a set of geographically close spatial units where the prevalence mean

Table II
CCR, MS AND *MAE* OF ALL THE DIFFERENT METHODS COMPARED
 WHEN THE CLASSIFIERS ARE APPLIED TO THE 2007 DATASET

	<i>CCR</i>	<i>MS</i>	<i>MAE</i>
KDLOR	80.30%	26.09%	0.2197
LDAOR	40.91%	0.00%	1.0076
C4.5	77.27%	0.00%	0.2803
SVM-Rank	79.54%	0.00%	0.2272
AdaBoost100	78.79%	34.78%	0.2348
MLP	75.76%	0.00%	0.2879

The best method is in bold face and the second best one in italics

Table III
 CONFUSION MATRICES OF ALL THE DIFFERENT METHODS COMPARED
 WHEN THE CLASSIFIERS ARE APPLIED TO THE 2007 DATASET

KDLOR				LDAOR			
Target	Predicted			Target	Predicted		
	#1	#2	#3		#1	#2	#3
#1	94	6	2	#1	48	0	54
#2	16	6	1	#2	10	0	13
#3	1	0	6	#3	1	0	6
C4.5				SVM-Rank			
Target	Predicted			Target	Predicted		
	#1	#2	#3		#1	#2	#3
#1	102	0	0	#1	99	1	2
#2	23	0	0	#2	22	0	1
#3	7	0	0	#3	1	0	6
AdaBoost100				MLP			
Target	Predicted			Target	Predicted		
	#1	#2	#3		#1	#2	#3
#1	90	10	2	#1	99	3	0
#2	14	8	1	#2	22	1	0
#3	1	0	6	#3	6	1	0

is significantly and uniformly —from a statistical point of view— high, showing this method a high performance. As mentioned, the method is based in the kernel version of the Linear Discriminant Analysis, taking into account the global information of classes, and using the optimal projection which maintain the rank of these classes. As shown, the algorithm improves the results obtained with C4.5, SVMRank, MLP and Adaboost.

In general, the KDLOR algorithm has demonstrated to be an useful and robust method to identify and locate hot and cold-spots for the selected illness. Using this ordinal approach, hot and cold-spots were identified and located in the space. It is well-known that merging tools for spatial analysis are needed in health technology assessment and spatial analysis, and the spatial distribution of depression in Andalusia at municipality level analysis shows that KDLOR is strategically relevant for health care management.

Since this example works with unbalanced classes, it could be interesting work with cost matrices but it is difficult to fix these because it is quite attached to the sample.

ACKNOWLEDGMENT

This work has been partially subsidized by the TIN 2008-06681-C06-03 project of the Spanish Ministerial Commission of Science and Technology (MICYT), FEDER funds

and the P08-TIC-3745 project of the “Junta de Andalucía” (Spain).

REFERENCES

- [1] B. Moreno, C. García-Alonso, M. A. Negrín-Hernandez, F. Torres-Gonzalez, and L. Salvador-Carulla, “Spatial analysis to identify hotspots of prevalence of schizophrenia,” *Social Psychiatry and Psychiatric Epidemiology*, vol. 43, no. 10, pp. 782–791, 2008.
- [2] B. Philips, G. Gong, K. Hargrave, E. Belasco, and C. Lyford, “Correlation of the ratio of metastatic to non-metastatic cancer cases with the degree of socioeconomic deprivation among texas counties,” *International Journal of Health Geographics*, vol. 10, no. 12, 2011.
- [3] C. Luke, B. Koczwara, C. Karapetis, K. Pittman, T. Price, D. Kotasek, K. Beckmann, M. P. Brown, and D. Roder, “Exploring the epidemiological characteristics of cancers of unknown primary site in an australian population: implications for research and clinical care,” *Australian and New Zealand J. of Public Health*, vol. 32, no. 4, pp. 383–9, 2008.
- [4] L. Pan, L. Qin, S. X. Yang, and J. Shuai, “A neural network-based method for risk factor analysis of west nile virus,” *Risk Analysis*, vol. 28, no. 2, pp. 487–96, 2008.
- [5] G. W. Dombi, J. P. Rosbolt, and R. K. Severson, “Neural network analysis of employment history as a risk factor for prostate cancer,” *Computers in Biology and Medicine*, vol. 40, pp. 751–757, 2010.
- [6] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 4th ed. Chapman & Hall/CRC, 2007.
- [7] L. Anselin, “Local indicators of spatial association-lisa,” *Geographical Analysis*, vol. 27, no. 2, pp. 93–115, 1995.
- [8] J. Wakefield, N. Best, and W. L., *Bayesian approaches to disease mapping. In Spatial epidemiology methods and applications*. Oxford: Oxford University Press, 2006.
- [9] C. A. Coello-Coello, G. B. Lamont, and D. A. van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [10] B.-Y. Sun, J. Li, D. D. Wu, X.-M. Zhang, and W.-B. Li, “Kernel discriminant learning for ordinal regression,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 906–910, 2010.
- [11] W. Chu and S. S. Keerthi, “Support vector ordinal regression,” *Neural Computation*, vol. 19, pp. 792–815, 2007.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed., ser. Inf. Science and Statistics. Springer, 2006.
- [13] M. A. B. T. Coleman and A. Grace, *Optimization Toolbox For use with Matlab - Users Guide*, 2nd ed., 1999.
- [14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The weka data mining software: an update,” *ACM SIGKDD Explor. Newsletter*, vol. 11, pp. 10–18, 2009.
- [15] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Elsevier, 2005.