DOI: 10.1515/bile-2016-0005

DE GRUYTER OPEN

Biometrical Letters
Vol. 53 (2016), No. 1, 37-46

Allocation of oaks to Kraft classes based on linear and nonlinear kernel discriminant variables

Bogna Zawieja[1], Katarzyna Kaźmierczak[2]

[1]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences. Wojska Polskiego 28, 60-637 Poznań, Poland, bogna13@up.poznan.pl
[2]Department of Forest Management, Poznań University of Life Sciences, Wojska Polskiego 71c, 60-625 Poznań, Poland, kkdendro@up.poznan.pl

SUMMARY

A method of discriminant variable determination was used to visualize the division of oak trees into Kraft classes. Usual discriminant variables and several types of kernel discriminant variables were studied. For this purpose the traits of oak (*Quercus* L.) trees, measured on standing trees, were used. These traits included height of tree, breast height diameter and crown projection area. The use of the Gaussian kernel and modified Gaussian kernel enabled the clearest division into Kraft classes. In particular, the latter method proved to be the most effective.

Key words: discriminant variables, kernel, Kraft classes, oak (*Quercus* L.)

1. Introduction

Researchers assessing the growth dynamics of trees deal with a single tree – its dimensions, age, increments, size of the crown and its elements, and growth space – when considering the properties of various species. A stand is not a collection of identical trees, but contains varied individuals which affect each other in many ways. Individual trees compete for light, growth space and available nutrients. The chances of individual trees are dependent on their growth energy and adaptive ability.

Attempts have been made to divide trees growing in a given area into groups using different classification systems. This has been done mainly for economic stand management, although some such systems can also be applied in other circumstances. The most commonly used classification system is that given by Kraft (1884), known as the "natural" system. This system is associated with the social position of trees as a combination of height relative to neighboring trees as well as crown size and quality. Kraft identified the following classes of trees: I – predominant trees, II – dominant trees, III – low co-dominant trees, IV – dominated trees (this class is divided into IVa – intermediate trees and IVb – partially overtopped crowns, the upper crown free, the lower crown under canopy cover) and V – entirely overtopped trees (divided into Va – with crowns capable of growth and Vb – with dead crowns). Classes I–III are called the dominant stand and classes IV–V the suppressed stand (see Zawieja and Kaźmierczak, 2015). Membership of a given social class reflects the position of a tree in a stand, and hence its growth potential. The assignment of trees to Kraft classes may be subjective.

Data can be presented graphically in order to show their configuration, grading into classes or outliers. However, if the data are multidimensional such presentation is very difficult (if three traits are observed) or even impossible (if more than three traits are observed). In this situation, discriminant variables can be used. In the study by Zawieja and Kaźmierczak (2015) ordinary discriminant variables and discriminant function analysis were used for the classification of Scots pine (*Pinus sylvestris* L.) trees according to several traits. This method does not always allow a clear separation among groups. In the present work several methods of graphical presentation of data obtained using usual discriminant variables and kernel discriminant variables are compared. The kernel discriminant variables were determined using a number of kernel functions including Gaussian kernel, Laplacian kernel and polynomial kernel. A modified Gaussian kernel was also used. The use of the latter method yielded the best separation of groups.

2. Experimental material

The empirical material was taken from two different stands. The first sample (stand A) was collected from a 0.25 ha sample plot, established in a 97-year-old oak stand in a mixed-deciduous forest site, with 94 oak trees growing within the plot. The second sample (stand B) was collected from a 0.75 ha sample plot, located in a 135-year-old oak stand in a moderately humid deciduous forest. The stand on the plot consisted of 160 trees in total, including 152 oaks and 8 pines.

The following traits of each tree were measured in both stands:

- $d_{1.3}$ – diameter at breast height (DBH) outside bark in two perpendicular directions (N-S and W-E) to the nearest 0.1 cm. The arithmetic mean of the two measures was treated as the real diameter of the tree;
- h – tree height (accuracy to 0.1 m);
- p_k – crown projection area, as a polygonal area with the projections of characteristic points of a crown: their locations were established with a mirror-based crown projector and the area was calculated using a polar method (distance from tree stump in cm and azimuth to 1°);
- l_k – height of live crown basis (accuracy to 0.1m).

3. Methods

Linear discriminant variables (Krzyśko 1990, 2000; Krzyśko et al. 2008; Kornacki and Ćwik 2005; Anderson 2003).

Let $\mathbf{X} = [x_{rij}]$ be a random sample from a normal distributed population divided into c ($r = 1,...,c$) groups, x_{rij} the observation of the i^{th} ($i = 1,...,n$) tree concerning the j^{th} ($j = 1,...,m$) trait in the r^{th} group, and $\mathbf{U}$ the $(t \times n)$ matrix of discriminant variables ($t = \min(c-1, m)$ is the number of discriminant variables). The discriminant variables $\mathbf{U}$ are presented as $\mathbf{U} = \mathbf{A'X}$, where $\mathbf{A'}$ is the ($t \times m$) matrix of coefficients of discriminant variables.

Let $\mathbf{x}_{ri} = [x_{rij}]$ be the m-vector of observations (one row of the matrix $\mathbf{X}$) of the i^{th} tree in the r^{th} group ($r = 1,...,c$), and $\bar{\mathbf{x}}_r = [\bar{x}_{rj}]$ the m-vector of means $\bar{x}_{rj} = n_r^{-1}\sum_{i=1}^{n_r} x_{rij}$ (where $\sum_{r=1}^{c} n_r = n$). Then $\bar{\mathbf{x}} = [\bar{x}_j]$ is the m-dimensional vector of means $\bar{x}_j = n^{-1}\sum_{r=1}^{c}\sum_{i=1}^{n_r} x_{rij}$. To determine $\mathbf{U}$ two matrices should be calculated. The first is the inter-class variation matrix $\mathbf{B} = \sum_{r=1}^{c} n_r(\bar{\mathbf{x}}_r - \bar{\mathbf{x}})(\bar{\mathbf{x}}_r - \bar{\mathbf{x}})'$ and the second is the within-class variation matrix $\mathbf{W} = \sum_{r=1}^{c}\sum_{i=1}^{n_r}(\mathbf{x}_{ri} - \bar{\mathbf{x}}_r)(\mathbf{x}_{ri} - \bar{\mathbf{x}}_r)'$. Let $\mathbf{S_B} = (c-1)^{-1}\mathbf{B}$ and $\mathbf{S_W} = (n-c)^{-1}\mathbf{W}$; then $\mathbf{S_T} = \mathbf{S_B} + \mathbf{S_W}$ is the matrix of total variation. The set of vectors $\mathbf{a}_l \in \mathrm{R}^m$ ($l = 1,...,t$; $t = \min(c-1, m)$; $\mathbf{a}_l$ is the l^{th} column of matrix $\mathbf{A}$) is sought for which the class differentiation index $J(\mathbf{a}) = \mathbf{a}'\mathbf{S_B}\mathbf{a}/\mathbf{a}'\mathbf{S_W}\mathbf{a}$ attains a maximum under the condition $\mathbf{a}_l'\mathbf{S_W}\mathbf{a}_k = \delta_{lk}$ (Kronecker delta). The vector which maximizes the index $J(\mathbf{a})$ is the eigenvector corresponding to the maximum eigenvalue of $\mathbf{S_W^{-1}}\mathbf{S_B}\mathbf{a}_l = \lambda_i\mathbf{a}_l$. Generally the solution of this equation is t non-negative eigenvalues λ_l ($\lambda_1 \geq \lambda_2 \geq ....\lambda_t \geq 0$) which correspond to eigenvectors $\mathbf{a}_l$. The first discriminant variable $\mathbf{a}_1'\mathbf{X}$ corresponds to the largest (first) eigenvalue, the second $\mathbf{a}_2'\mathbf{X}$ to the second eigenvalue, etc.

The discriminant variable $\mathbf{u}_l = \mathbf{a}_l'\mathbf{X}$ is useful in the classification process if the eigenvalue λ_l is significantly different from zero. An unknown number of non-zero eigenvalues $d = 0,1,...,t-1$ are determined by sequential testing of hypotheses. In the first step the hypothesis that all eigenvalues are equal to zero is tested. If it is rejected, in the second step the hypothesis that $t-1$ eigenvalues are equal to zero (ignoring the largest eigenvalue λ_1) is tested, and so on. This process is continued until the l^{th} hypothesis is not rejected, since then the remaining eigenvalues are equal to zero ($d = l$). Wilks' lambda statistic $\Lambda_d = \prod_{l=d+1}^{t} 1/(1+\lambda_l)$ is used to verify these hypotheses. The $\chi_B^2 = -[n-c-(m-c+2)/2]\ln\Lambda_d$ have an asymptotic χ^2 distribution with $(m-d)\cdot(c-d-1)$ degrees of freedom.

Nonlinear kernel discriminant variables (Mika et al. 1999; Baudat and Anoura 2000; Deręgowski and Krzyśko 2014).

The vector of standardized observations $\mathbf{x}_i = (x_{11}, x_{12},..., x_{1m})'$ belongs to the space R^m, which is transformed nonlinearly in Hilbert space with reproducing kernel. The transformed function is the vector function $\phi : \mathrm{R}^m \to \mathcal{H}_k$; this nonlinear transformation $\phi(\mathbf{x}_i) = \tilde{\mathbf{x}}_i$ is not known. However, only the known form of the non-negative definite kernel function $k(\mathbf{x}_i, \mathbf{x}_{i'}) = \phi'(\mathbf{x}_i)\phi(\mathbf{x}_{i'})$ is needed for the selection. Let $\tilde{\mathbf{B}} = \sum_{r=1}^{c} n_r (\tilde{\mathbf{x}}_r - \tilde{\mathbf{x}})(\tilde{\mathbf{x}}_r - \tilde{\mathbf{x}})'$ be the inter-class variation matrix ($\tilde{\mathbf{x}}_r$ is the vector of means of transformed observations in the r^{th} class, $\tilde{\mathbf{x}}$ is the vector of total means of transformed observations), let $\tilde{\mathbf{W}} = \sum_{r=1}^{c} \sum_{i=1}^{n_r} (\tilde{\mathbf{x}}_{ri} - \tilde{\mathbf{x}}_r)(\tilde{\mathbf{x}}_{ri} - \tilde{\mathbf{x}}_r)'$ be the within-class variation matrix ($\tilde{\mathbf{x}}_{ri}$ is the vector of transformed observations), and let $\tilde{\mathbf{S}}_{\mathbf{B}} = (c-1)^{-1}\tilde{\mathbf{B}}$ and $\tilde{\mathbf{S}}_{\mathbf{W}} = (n-c)^{-1}\tilde{\mathbf{W}}$; then $\tilde{\mathbf{S}}_{\mathbf{T}} = \tilde{\mathbf{S}}_{\mathbf{B}} + \tilde{\mathbf{S}}_{\mathbf{W}}$ is the matrix of total variation. Finding discriminatory variables in the trait space is reduced to solving the optimization problem $\mathbf{b}_o = \arg\max(\mathbf{b}'\tilde{\mathbf{S}}_{\mathbf{B}}\mathbf{b} / \mathbf{b}'\tilde{\mathbf{S}}_{\mathbf{W}}\mathbf{b})$. This optimization problem is equivalent to $\mathbf{b}_o = \arg\max(\mathbf{b}'\mathbf{KDKb} / \mathbf{b}'\mathbf{KKb})$, where the elements of the matrix $\mathbf{D}$ are $1/n_r$ if $\mathbf{x}_i$ and $\mathbf{x}_{i'}$ belong to the same class, and zero otherwise. The vectors $\mathbf{b}_o$ are equal eigenvectors corresponding to the maximum eigenvalues in the generalized eigenproblem $\mathbf{KDKb}_l = \lambda_l \mathbf{KKb}_l$, where $\mathbf{K} = \mathbf{P}\tilde{\mathbf{K}}\mathbf{P}$ with $\mathbf{P} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n$ (matrix of centering) and $\tilde{\mathbf{K}} = (k_{ii'})$ with $k_{ii'} = k(\mathbf{x}_i, \mathbf{x}_{i'})$ is the kernel matrix. Since both of the matrices $\mathbf{KDK}$ and $\mathbf{KK}$ are non-negative definite, for the purpose of solving the above eigenproblem the inverse of the matrix ($\mathbf{KK} + \varepsilon\mathbf{I}$) is used. Then the eigenvalues and eigenvectors of the matrix $(\mathbf{KK} + \varepsilon\mathbf{I})^{-1}\mathbf{KDK}$ are solutions to the above optimization problem. The kernel discriminant variables are in the form $\mathbf{v}_l = \mathbf{b}_l'\mathbf{K}$.

The four kernel functions are used for the calculation of nonlinear discriminant variables. Three of them are the ordinary kernel functions, including the Laplacian kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\| / \sigma)$ (for $\sigma > 0$), the homogeneous polynomial kernel $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}'\mathbf{y})^d$ and the Gaussian kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / \sigma)$. Moreover the modified kernel function is used in the form $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^4 / \sigma)$. In this study the values $d = 1$ (which means that the polynomial kernel is identical to the linear kernel) and $\sigma^{-1} = 0.00004$ were used. The values of these parameters are determined in order to obtain the best image of the data.

4. Results

As a result of the use of linear discriminant variables, the analyzed stands could be divided visually into three classes (Figure 1) in the case of stand A and two classes in the case of stand B. In both stands, class IV stood out from the remainder. In Table 1 the eigenvalues, significance of Wilks' lambda statistics and participation of discriminant variables in the total variability are given. Only the two first discriminant variables were statistically significant; thus the data can be presented in the coordinate system of these variables.

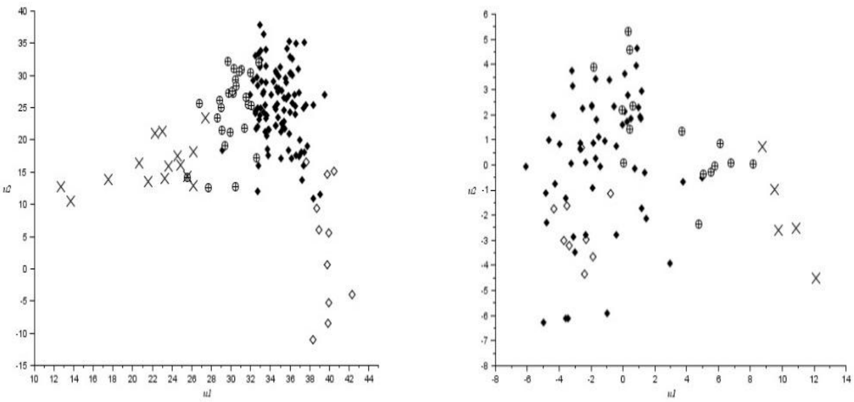


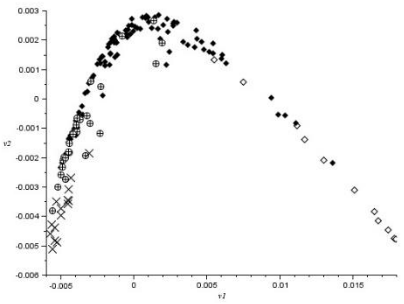
Figure 1. Stand A (left panel) and B (right panel) in the coordinate system of the first two linear discriminant variables (empty diamonds – Kraft class I, full diamonds – Kraft class II, circles with a cross – Kraft class III, diagonal crosses – Kraft class IV)

Table 1. Parameters of linear discriminant variables

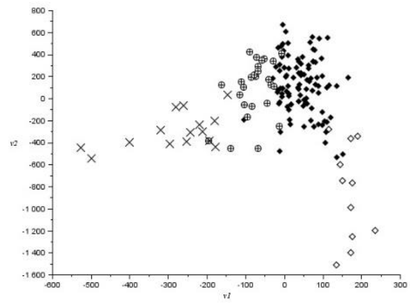
	stand A			stand B		
u_l	eigenvalues	*p*-value	participation %)	eigenvalues	*p*-value	participation (%)
1	4.0048	<0.0001	79.23	1.8215	<0.0001	88.81
2	1.0351	<0.0001	20.48	0.1972	0.0099	9.62
3	0.0145	0.3461	0.29	0.0323	0.2823	1.58

p-value of Wilks' lambda statistic

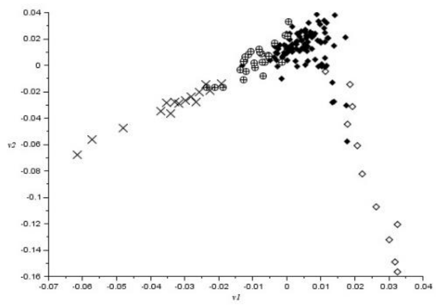
The results of the application of nonlinear kernel discriminant variables are presented in Figures 2 and 3 for stands A and B respectively. The use of three ordinal kernels resulted in the distinguishing of three groups of trees in the case of stand A and two groups in the case of stand B. These results are quite similar to the results obtained with the use of linear discriminant variables. The application of the modified kernel function resulted in a clear division into four groups of trees in the case of stand A, and also, with perhaps a somewhat less clear division, in the case of stand B.



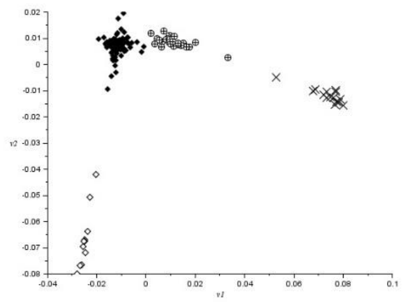
Laplacian kernel



Homogenous polynomial kernel

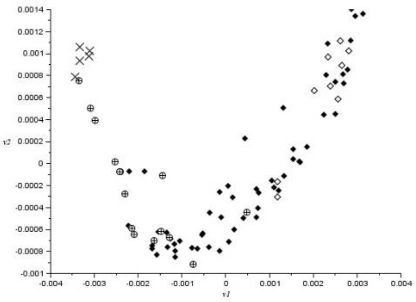


Gaussian kernel

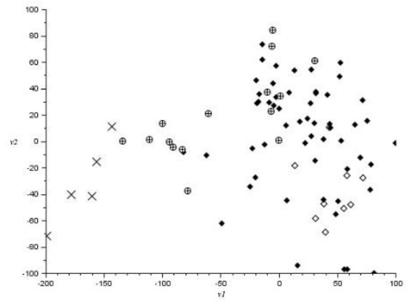


Modified kernel

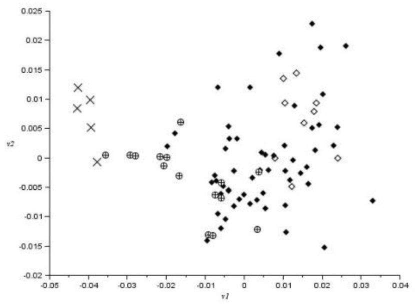
Figure 2. Stand A in the system of the first two nonlinear discriminant variables



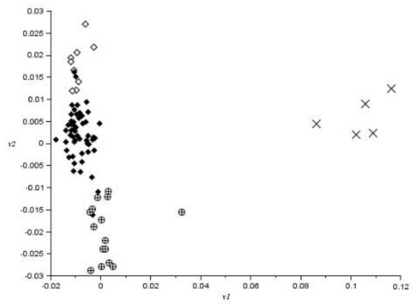
Laplacian kernel



Homogenous polynomial kernel



Gaussian kernel



Modified kernel

Figure 3. Stand B in the system of the first two nonlinear discriminant variables

5. Discussion and conclusions

The Kraft classification system describes the social position of trees in terms of a combination of height relative to neighboring trees as well as crown size and quality (Kraft 1884). The classification of trees may be subjective. However, the total proportion of trees in dominating and dominated stands corresponds well with their participation in total volume production (Jaworski 2004). The dominating stand comprises about 70% of trees, their basal area accounts for 80% of the total basal area of the stand, while their thickwood volume accounts for almost 90% (Korpel after Jaworski, 2004). The biosocial position of trees has been found to have a significant influence on increments of Scots pine trees (Kaźmierczak 2013). Position in the stand is highly correlated with measured traits of trees (Kaźmierczak 2009, 2010, 2012).

An attempt to visualize the grading into Kraft classes of Scots pine trees was made by Zawieja and Kaźmierczak (2015). In that study, measurable traits such as tree height, breast height diameter inside bark, tree basal area, double bark thickness, and tree slenderness defined as the ratio of height to breast height diameter were used to divide trees into groups using linear discriminant analysis. The traits that were found to give the best division into groups were height, ten-year (or five-year) increments at breast height, slenderness, and breast height diameter without bark. In a paper by Grala-Michalak and Kaźmierczak (2011) the similar problem of allocation of trees to Kraft classes was considered. Here, traits that can be measured on fallen trees were also considered in the discrimination analysis.

Linear discriminant analysis was used to distinguish floristically different forest types by Thessler et al. (2008). Apart from discriminant analysis, those authors applied the non-parametric *k* nearest neighbors (k-nn) classifier to distinguish classes. In turn, the use of Fisher discriminant analysis was demonstrated by Jing et al. (2015), where based on multivariate statistical analysis the status of forest fire risk points was assessed.

In the present work the use of kernel nonlinear discriminant variables made it possible to obtain a clearer visual division into groups, especially when the Gaussian and modified kernels were used. The latter gave a much clearer picture.

REFERENCES

Anderson T.W. (2003): An introduction to multivariate statistical analysis. New York, Wiley.

Baudat G., Anour F. (2000): Generalized discriminant analysis using a kernel approach. Neural Computation 12: 2385-2404.

Dergowski K., Krzyśko M. (2014): A kernel based learning algorithm combining kernel discriminant coordinates and kernel principal components. Biometrical Letters 51(1): 57-73.

Grala-Michalak J., Kaźmierczak K. (2011): Discriminant analysis for Kraft's classes of trees. Biometrical Letters 48(1): 67-81.

Jaworski A. (2004): Fundamentals of incremental and ecological regeneration and stands tending (in Polish). Państwowe Wydawnictwo Rolnicze i Leśne, Warsaw, Poland.

Jing Z., Weiqing M., Ye Z. (2015): Fisher Linear Discriminant Method for Forest Fire Risk Points on Transmission Line. International Journal of Smart Home. 9(4): 25-34. http://dx.doi.org/10.14257/ijsh.2015.9.4.03

Kaźmierczak K. (2009): Selected measures of the growth space of a single tree in maturing pine stand (in Polish). Sylwan 5: 298-303.

Kaźmierczak K. (2010): Selected measures of the growth space of a single tree in a 50-year-old Scots pine stand (in Polish). Sylwan 154(4): 267-274.

Kaźmierczak K. (2012): Selected measures of the growth space of a single tree in a 35-years-old pine stand (in Polish). Sylwan 156(4): 280-286.

Kaźmierczak K. (2013): The current growth increment of pine tree stands comprising three different age classes (in Polish). Leśne Prace Badawcze, 74(2): 93-100.

Kornacki J., Ćwik J. (2005): Statistical learning systems (in Polish). WNT Warsaw.

Kraft G. (1884): Beiträge zur Lehre von den Durchforstungen, Schlagstellungen und Lichtungshieben. Hannover, Klindworth.

Krzyśko M. (1990): Discriminant analysis (in Polish). WNT, Warsaw.

Krzyśko M. (2009): Fundamentals of multivariate statistical inference (in Polish). Wydawnictwo Naukowe UAM, Poznań: 267-278.

Krzyśko M., Wołyński W., Górecki T., Skrzybut M. (2008): Learning systems. Pattern Recognition. Cluster analysis, reduction of dimensionality (in Polish). WNT Warsaw.

Mika S., Rätsch G., Weston J., Schölkopf B., Müller K.R. (1999): Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing, Hu Y.H., Larsen J., Wilson E., and Douglas S. (eds.), IX: 41-48.

Thessler S., Sesnie S., Ramos Bendaña Z.S., Ruokolainen K., Tomppo E., Finegan B. (2008): Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica. Remote Sensing of Environment, 112(5): 2485-2494.

Zawieja B., Kaźmierczak K. (2015): The method of standing trees allocation to different biosocial classes. Colloquium Biometricum 45: 78-92.