Spatial Prediction with Categorical Response Variables

Martin Charlton[*1], Chris Brunsdon[†1], Paul Harris[‡2] and Lex Comber[§3]

[1]National Centre for Geocomputation, Maynooth University, Maynooth, Co Kildare, Ireland
[2]Rothamsted Research, North Wyke, Okehampton, Devon EX20 2SB
[3]School of Geography, University of Leeds, Leeds LS2 9JT

December 3, 2017

Summary

Discriminant analysis (DA) is a useful multivariate tool when the response variable is categorical. While logistic regression may be used when the response variable has two categories, discriminant analysis can be appropriate when there are more than two categories. We outline an extension of linear and quadratic DA to allow for geographical weighting. We discuss how the output from GDWA may be displayed, including the use of geographically weighted crosstabulations. An example dataset extracted from the FAO Global Land Cover map is used to illustrate GWDA in practice.

KEYWORDS: geographically weighted discriminant analysis, spatial prediction.

1. Introduction

There are occasions when it is desired to predict an object's membership of a particular discrete group. The response variable may have two or more values. There are a number of possible techniques which are available. When the response variable takes on only two values, say 0 and 1, the analyst may choose logistic regression. The goal of logistic regression is to predict the probability of membership of a defined group (usually that with the value 1). Logistic regression may also be used with a multinomial response variable (Hosmer and Lemeshow, 1989).

An alternative technique is provided by discriminant function analysis (Fisher, 1936), as generalised by Rao (1948) which allows for the extraction of discriminant functions from the independent variables which are then used to generated probabilities of membership of the groups in question for each observation. If there are k groups in the dataset, indexed by the dependent variable, the goal of the technique is to extract k discriminant functions. Marks and Dunn (1974) consider the extension of the linear classifier to the quadratic case. An observation is assigned to group j is the value for the discriminant function for the group is the smallest.

We assume that the data are multivariate normal. If $\boldsymbol{\Sigma}_j$ is the variance-covariance matrix for the members of group j, q is the number of predictor variables in $\mathbf{x}$, μ_j is the mean vector for the observations in group j, and p_j is the prior probability of membership of group j, the linear assignment rule (LDA) can be written as

[*] martin.charlton@mu.ie
[†] christopher.brunsdon@mu.ie
[‡] paul.harris@rothamsted.ac.uk
[§] a.comber@leeds.ac.uk

$$k = \arg\min_{j\in(1,\ldots,m)} \left[\frac{q}{2}\log(|\Sigma|) + \frac{1}{2}(x-\mu_j)^T \Sigma^{-1}(x-\mu_j) - \log(p_j) \right] \tag{1}$$

and the quadratic assignment rule (QDA) can be written

$$k = \arg\min_{j\in(1,\ldots,m)} \left[\frac{q}{2}\log(|\Sigma_j|) - x^T\Sigma_j^{-1}\mu_j + \frac{1}{2}\mu_j^T\Sigma_j^{-1}\mu_j - \log(p_j) \right] \tag{2}$$

Marks and Dunn (1974) and Wahl and Kronmal (1977) examine the behaviour of these functions with data where the covariances are unequal. For small samples the linear function appears to provide more reliable assignment, whereas for large samples, the quadratic function is to be preferred. However, if the covariance differences are substantial neither rule may provide reliable classification.

The reliability of the classification can be explored in a number of ways. If a correspondence or confusion matrix is generated by crosstabulating the observed classes (rows) with the predicted classes (columns) the proportion of correctly classified observations is the ratio of the trace of the matrix to the total number of observations. This ratio is also known as the *portmanteau accuracy* or *user's accuracy*.

A slightly different measure of the reliability is provided by the *producer's accuracy* – the ratio of the number of correctly classified objects in each class to the total number of objects in that predicted class. This is sometimes known as *kappa*.

2. Geographical Weighting

By analogy with geographically weighted regression and geographically weighted principal components analysis, the variance covariance matrices, mean vectors, and prior probabilities can be geographically weighted relative to some arbitrary location in the study area ***u***. Details of this are available in Brunsdon et al (2007).

In the implementation in the GWModel library (Lu *et al*, 2014), the functions for fitting geographically weighted discriminant models include arguments to specify global or local covariances, means and priors, and linear or quadratic models. The defaults are for local weighting, linear assignment, and a fixed radius bisquare kernel (as with the other GW methods). Separate training and validation datasets can also be specified.

As with unweighted LDA and QDA, the output includes a vector of the assignments after the corresponding rule has been applied. The values of the discriminant functions for each group in the observed data are also reported for each observation. These can be converted to the posterior probabilities. With the posterior probabilities we can then compute the relative entropy of the probabilities.

There are a series of mapping choices – we can map the assigned group, perhaps with a symbol to indicate whether the assignment has been correct or not. Mapping the probability associated with the assignment shows us where we can be more, or less, sure of the prediction. An observation which is misclassified, but has a high probability, might be regarded as an outlier, and worthy of further investigation. Mapping the spatial variation in the entropies of the posterior probabilities can reveal where there is ambiguity in the choice of the predicted class. If the entropy is low, then there is less ambiguity in the assignment than if the entropy is high (the entropy will be highest is the posterior probabilities are equal). Again, areas of high entropy might be worthy of further investigation. Brunsdon et al (2007) investigate using colour variations in the CIE colour chart to visualise spatial ambiguity but this is restricted to 3 classes.

3. Geographically weighted crosstabulations

A second approach is explored by Comber et al (2017). The correspondence matrix for the assignments can be locally weighted, and a variety of measures based on them mapped as well. These include locally weighted producer's and user's accuracy measures. One might also map the locally weighted χ^2 statistic: this will be larger if there is a stronger diagonality in the tables (when the assignment is more reliable); a low values will suggest poor local assignment.

4. Experiment: European land cover imagery

We use the methods outlined above to explore that relationship between the land cover classes in the FAO 2013 global land cover (GLC) map, and the predictors outlined in Table 1.

Table 1 Predictor variables

Variable
Hansen tree cover loss 2000-2013 (percentage of the pixel
DMSP-OLS night-time lights N 2010
Human impact assessment from the IIASA Geo-Wiki project (0-100)
IIASA cropland map (percent of the pixel)

The FAO data consists of 11 classes (different from the 10 used in Comber et al (2017)), of which the most numerous were [2] cropland, [3] grassland and [4] tree-covered areas. The definitions are given in Annex 1 of Latham et al (2014). The goal of the exercise is to predict the membership of the land cover class using 2766 locations identified in the IIASA Geo-Wiki. The data were subdivided into a training set of 1697 observations and a validation set of 682 observations (some records were incomplete and therefore omitted).

Following the practice with other geographically weighted approaches, analysis should begin with suitable data exploration. Boxplots are helpful in assessing the distribution of the predictor variables within the 3 land cover classes we have chosen. These are shown in Figure 1.

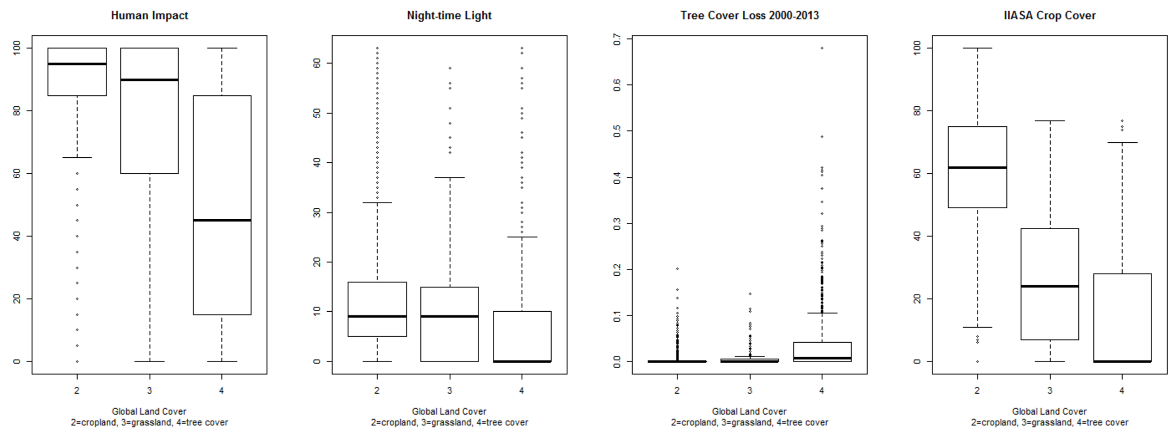

Figure 1 Within and between class differences

The skewed nature of the distributions within each class should be noted. The zero values in the light and tree cover data preclude the application of a log transform, although a square root transformation may tame some of the long right tails.

We proceed by fitting global models, and examining the various accuracy measures. It would be useful to look at the influence of the IIASA cropland variable on the accuracy of the predictions. As with the local models, mapping the spatial variation in the posterior probability of the winning class, and the entropy of the posteriors can be revealing.

In each case the models are fitted to the training dataset, and the accuracy and mapping applied to the data from the validation set.

The GWDA analysis proceeds along similar lines. The bandwidth is calibrated to maximise the portmanteau accuracy using the training data, and then the predictions of the posterior probabilities are carried out using the validation data.

A characteristic of the output from the spatial data frame in the gwda function in GWmodel is that the outputs are the values of the discriminant functions themselves (not posterior probabilities) and they are represented as character factors. Before the posterior probabilities can be computed, the relevant columns are converted to numeric, exponentials taken, and the rows normalised to sum to unity. Following this the maximum probability is identified, and the relative Shannon entropy computed for each row of posterior probabilities. We have some R code to carry these computations out, and this will eventually appear in GWmodel. The posterior probabilities thus computed, with a bandwidth of 100000km, match those from the global model (the functions *lda()* and *qda()* are in the MASS library.

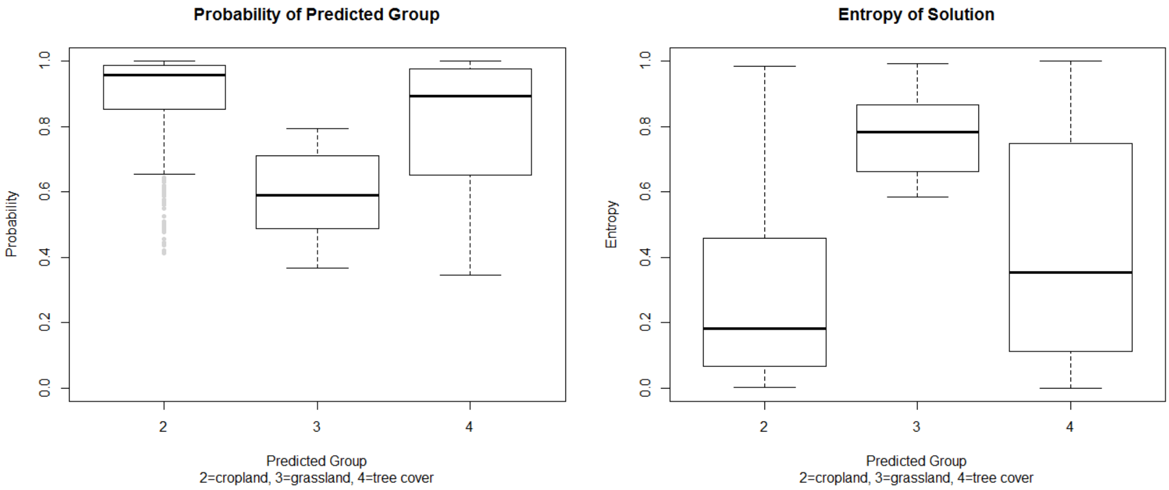

Figure 2 Certainty and ambiguity from GWDA

Whilst the global predictive behaviour is satisfying, the local predictive capability is slightly higher. Part of the improvement might be due the inclusion of the spatial effects in the model. Group 3, grassland, is least well predicted. The distributions of probabilities for cropland and tree cover are mirrored by the lower choice ambiguity revealed by the entropy measures.

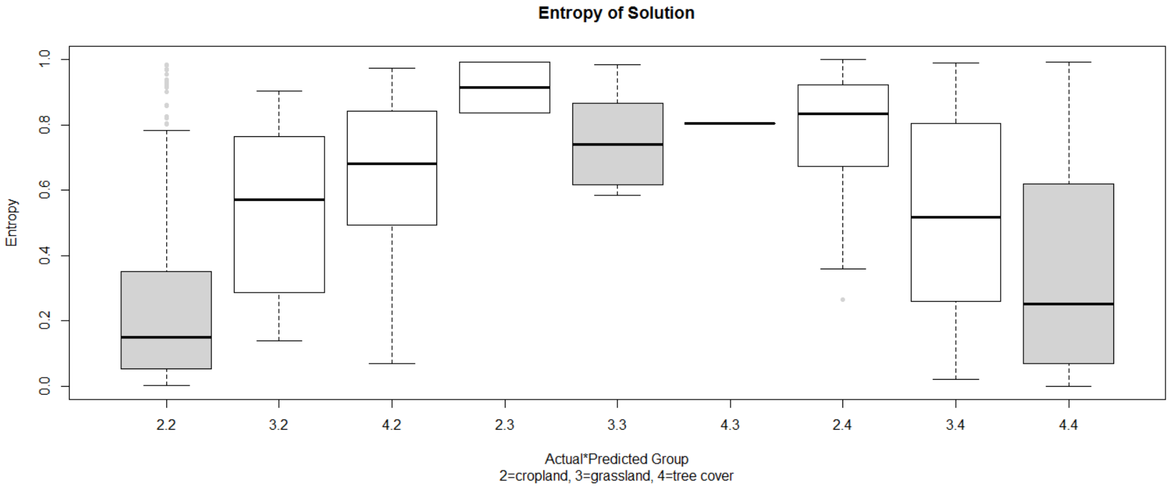

Figure 3 Cross-classified ambiguity measure

The boxplots of the ambiguity measures cross-classified by the observed and predicted classes, shown in Figure 3, are revealing.

The grey boxplot in each group of three is the correctly classified class; the median value is lowest in each group. For the class 3 the ambiguities are high and remain obstinately high; grassland is a challenge to predict given the covariates we have employed.

We may also map the variation in uncertainty by comparing the global and local patterns of classification success through the geographically weighted portmanteau accuracy and kappa. Because we cannot show both on a plot, we choose symbols based on the comparison with the global value. In Figure 4, locations where both locally weighted measures exceed the global measure, then symbol is +, for the converse case it is -, and where only one is higher, the symbol is *.

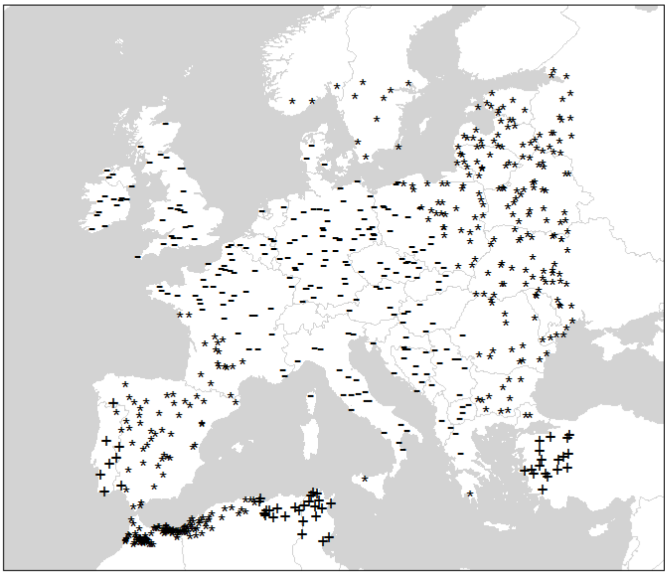

Figure 4 GW crosstab of portmanteau accuracy and kappa

The pattern is interesting, reflecting the strength of the areas where classification has been particularly successful: NE Algeria, N Tunisia, and W Turkey. A NW-SE band of less impressive levels of performance runs from the UK and Ireland, through France and Italy in the west to Poland the W Balkans in the east.

5. Conclusions

At the moment we are reporting work in progress. While GW logistical and binomial regression has existed for some time, we have been slow in using GWDA. It does provide some interesting challenges in visualisation, but there are plenty of possibilities and challenges. Examination of the discriminant function loadings, a future development, might allow us to understand which variables are locally influencing misclassification.

6. Acknowledgements

We gratefully acknowledge support from Science Foundation Ireland's Strategic Research Initiative (07/SRC/I1168) which allowed some of the development of geographically weighted discriminant methods. We also acknowledge support from Science Foundation Ireland's Investigator Award Programme (15/IA/3090). The data for the experiments was kindly made available by Linda See and Steffen Fritz, and prepared by Myroslava Lesiv and Olga Danylo, from the International Institute for Applied Systems Analysis, Laxenburg, Austria.

7. Biography

Martin Charlton is Senior Lecturer in Geocomputation at Maynooth University. He is often to be found in the Roost.

Chris Brunsdon is Professor of Geocomputation at Maynooth University. He is also a regular in the Roost.

Paul Harris is Project Leader in Sustainable Agriculture Sciences at Rothamsted Research's North Wyke Farm Platform in Devon. He used to work at Maynooth University. He knows the Roost well.

Lex Comber is Professor of Geography at the University of Leeds. He wrote the abstract of a paper for GIScience 2016 on a flight from Rhodes to London, aided by Martin and Chris. And he knows the Roost.

References

Brunsdon C, Fotheringham S, and Charlton M (2007). Geographically weighted discriminant analysis. *Geographical Analysis*, 39, 376-396

Comber L, Brunsdon C, Charlton M and Harris P (2017). Geographically weighted correspondence matrices for local error reporting and change analyses: mapping the spatial distribution of errors and change. *Remote Sensing Letters*, 8(3), 234-243

Fisher R A (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* 7(2), 179-188.

Hosmer D W and Lemeshow S (1995). *Applied logistic regression*. Wiley, New York.

Latham J, Cumai R, Rosati I and Bloise M (2014). *Global Land Cover SHARE database Beta-Release Version 1.0 – 2014,* Food and Agriculture Organisation of the United Nations

Lu B, Harris P, Charlton M, and Brunsdon C (2014). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science,* 17(2), 85-101

Marks S and Dunn O J (1974). Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association*, 69, 555-559

Rao C R (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B*, 10(2), 159-203

Wahl P and Kronmal RA (1977). Discriminant functions when covariances are unequal and sample sizes are moderate. *Biometrics*, 33(3), 479-484