P-values for classification

Lutz Dümbgen, *University of Berne*
Bernd-Wolfgang Igl, *University at Luebeck*
Axel Munk, *University of Goettingen*

Abstract

Let (X,Y) be a random variable consisting of an observed feature vector $X in XX$ and an unobserved class label $Y in {1,2,ldots,L}$ with unknown joint distribution. In addition, let DD be a training data set consisting of n completely observed independent copies of (X,Y). Usual classification procedures provide point predictors (classifiers) $hat{Y}(X,DD)$ of Y or estimate the conditional distribution of Y given X. In order to quantify the certainty of classifying X we propose to construct for each $theta = 1,2,ldots,L$ a p-value $pi_theta(X,DD)$ for the null hypothesis that $Y = theta$, treating Y temporarily as a fixed parameter. In other words, the point predictor $hat{Y}(X,DD)$ is replaced with a prediction region for Y with a certain confidence. We argue that (i) this approach is advantageous over traditional approaches and (ii) any reasonable classifier can be modified to yield nonparametric p-values. We discuss issues such as optimality, single use and multiple use validity, as well as computational and graphical aspects.

AMS 2000 subject classifications: 62C05, 62F25, 62G09, 62G15, 62H30.

Keywords: nearest neighbors, nonparametric, optimality, permutation test, prediction region, ROC curve, typicality index, validity.

Full Text: PDF

Dümbgen, Lutz, Igl, Bernd-Wolfgang, Munk, Axel, P-values for classification, Electronic Journal of Statistics, 2, (2008), 468-493 (electronic). DOI: 10.1214/08-EJS245.

References

[1] Ehm, W., E. Mammen and D.W. Müller (1995). Power robustification of approximately linear tests. J. Amer. Statist. Assoc. 90, 1025–1033. MR1354019

[2] Federer, H. (1969). Geometric Measure Theory. Springer, Berlin Heidelberg. MR0257325

[3] Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. Ann. Eugenics 7, 179–184.

[4] Fraley, C. and A.E. Raftery (2002). Model-based clustering, discriminant analysis and density estimation. J. Amer. Statist. Assoc. 97, 611–631. MR1951635

[5] Holzmann, H., A. Munk and B. Stratmann (2004). Identifiability of finite mixtures - with applications to circular distributions. Sankhya 66, 440–450. MR2108200

[6] Holzmann, H., A. Munk and T. Gneiting (2006). Identifiability of finite mixtures of elliptical distributions. Scand. J. Statist. 33, 753-763. MR2300914

[7] McLachlan, G.J. (1992). Discriminant Analysis and Statistical Pattern Recognition. Wiley, New York. MR1190469

[8] Peel, D. and G.J. McLachlan (2000). Robust mixture modeling using the t-distribution. Statist. Computing 10, 339–348.

[9] Peel, D., W.J. Whitten and G.J. McLachlan (2001). Fitting mixtures of Kent distributions to aid in joint set identification. J. Amer. Statist. Assoc. 96, 56–63. MR1973782

[10] Ripley, B.D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge, UK. MR1438788

[11] Shorack, G.R. and J.A. Wellner (1986). Empirical Processes with Applications to Statistics. Wiley, New York. MR0838963

[12] Stone, C.J. (1977). Consistent nonparametric regression. Ann. Statist. 5, 595–645. MR0443204

[13] Yakowitz, S.J. and J.D. Spragins (1968). On the identifiability of finite mixtures. Ann. Math. Statist. 39, 209–214. MR0224204

Electronic Journal of Statistics. ISSN: 1935-7524