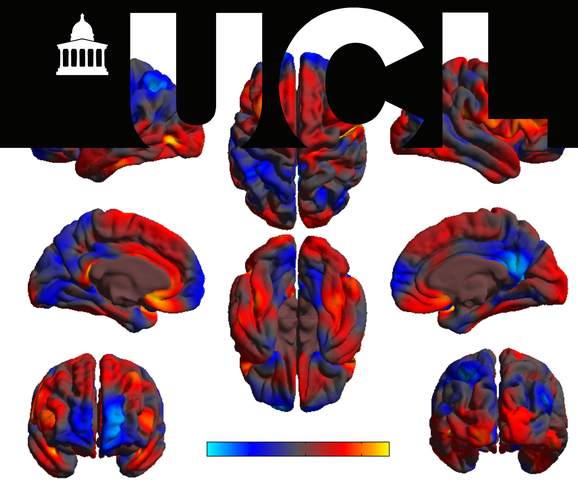
Balanced sensitivity and specificity on unbalanced data using support vector machine re-thresholding

Gerard Ridgway, Manja Lehmann, Jonathan Rohrer, Matthew Clarkson, Jason Warren, Sébastien Ourselin, Nick Fox
Dementia Research Centre, UCL Institute of Neurology, Queen Square, London, UK



Summary

Support vector machine (SVM) classifiers use multivariate patterns to separate two groups by a hyperplane with maximal margin, as shown in Figure 1. This strategy tends to obtain good generalisation accuracy on even very high dimensional applications. However, SVMs are not well suited to unbalanced data with very different numbers of cases in each group. In this work we implement a properly cross-validated method for altering the SVM threshold (also known as the bias or cut-point) to re-balance the sensitivity and specificity.

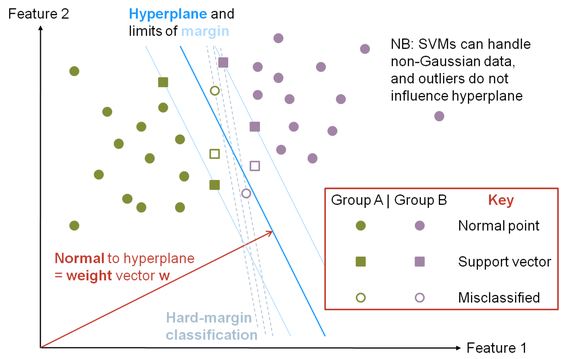


Figure 1: Illustration of the SVM for a two-dimensional example. The features could be things like mean cortical thickness over two regions of interest, or the values at two vertices on the triangulated surface. The hyperplane can be defined by $\mathbf{w}^T\mathbf{x} + b = 0$, which generalises to arbitrary length feature vectors $\mathbf{x} = [\text{Feature}_1, \text{Feature}_2, \ldots]$. For linear SVMs, the weight vector $\mathbf{w}$ has the same dimension as the feature vector, and hence can be visualised as an image, as in the graphic behind the UCL logo.

Introduction

Machine learning approaches for classifying individuals into disease groups on the basis of neuroimaging data are increasingly popular, and often use SVMs to handle dimensionalities much larger than the number of cases. SVMs have been applied to discriminate AD from healthy ageing, and predict MCI progression to AD. Their key motivation is that characterisation of multivariate patterns in the data should be more powerful than simpler mass-univariate analyses.

However, SVMs are intended for balanced data-sets with roughly equal numbers of positive and negative cases, aiming to maximise accuracy. With unbalanced data, sensitivity and specificity can be strongly biased in opposite directions. Consideration of receiver operating characteristics (ROC) can be used to estimate balanced sensitivity and specificity, but these may be upwardly biased due to their post-hoc nature. Unfortunately, many data-sets are unbalanced, for example from natural imbalances in prevalence of diseases or genetic variants, or low proportions of MCI subjects progressing over short periods.

Methods

We use nested leave-one-out cross-validation (CV), in which the SVM's parameters (including the soft-margin misclassification cost C) are trained within the inner CV loop. Here, the SVM's threshold or bias b can also be altered to re-balance sensitivity and specificity, by considering the ROC on the inner CV loop. Various criteria could be used to define the optimal threshold, such as highest Youden index (the sum of sensitivity and specificity); experiments suggest the best performance is obtained by directly minimising the difference between sensitivity and specificity, which is the criterion used here.

We evaluate the method by revisiting an unbalanced data-set of 34 subjects who presented clinically with frontotemporal dementia, of whom 23 exhibited frontotemporal lobar degeneration pathology (FTLD-FTD) and 11 had underlying AD pathology (AD-FTD) at post mortem (Lehmann et al., 2010). The small number of subjects here reflects the rarity of subjects who have both in vivo MRI scans and post mortem histopathology. The smaller number of unusual AD-FTD subjects is also practically unavoidable.

The features used in the SVM are the cortical thickness measurements (obtained from FreeSurfer, version 4.03, `http://surfer.nmr.mgh.harvard.edu`) at every vertex across both hemispheres of the cortex, with a dimensionality of 327,684.

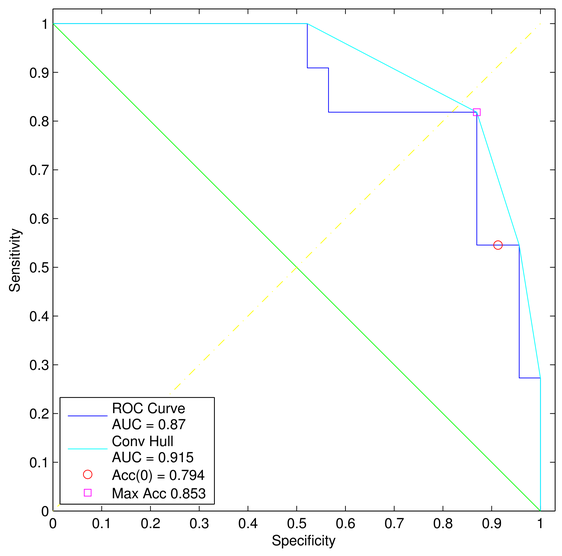


Figure 2: A posteriori receiver operating characteristic (ROC) curve for the original SVM, showing the sensitivity to the 11 AD-FTD subjects and the specificity to the 23 FTLD-FTD subjects. The red circle indicates the a priori operating point as learnt by the SVM, which has very poor sensitivity to the minority class. The square indicates the highest a posteriori accuracy, which in this case also achieves nearly equal sensitivity and specificity. However, simulations using random data show that the a posteriori accuracy is upwardly biased.

Results

In Lehmann et al. (2010), an SVM applied to the same data yielded an ROC curve encompassing a large area, but whose a priori sensitivity to the smaller AD-FTD group was near chance (see Figure 2 and Table 1). The new re-thresholded SVM attained high sensitivity and specificity without post-hoc modification (shown in Figure 3 and the second row of Table 1).

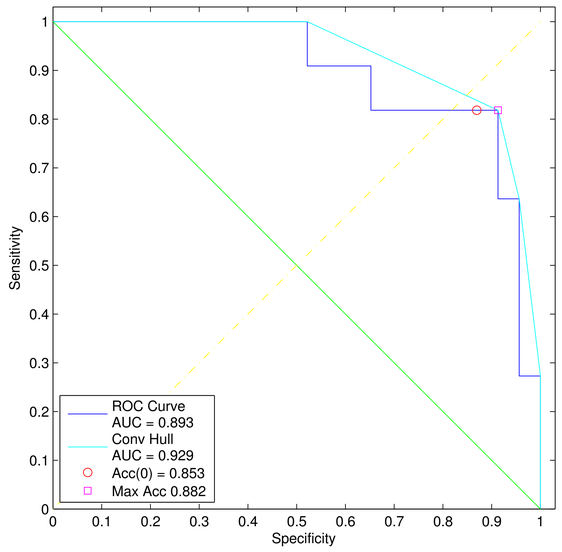


Figure 3: ROC curve for the modified (re-thresholded) SVM, showing that the a priori operating point (red circle) is now near optimal.

Method	Accuracy % (CI)	Sensitivity % (CI)	Specificity % (CI)
Original	79.4 (62.1–91.3)	54.5 (23.4–83.3)	91.3 (72.0–98.9)
Modified	82.4 (65.5–93.2)	81.8 (48.2–97.7)	82.6 (61.2–95.0)

Table 1: Classification performance metrics, without post-hoc alteration of cut-points. Binomial (exact) 95% confidence intervals are given in parentheses for each measure. The accuracy is not significantly different, but the sensitivity has been substantially increased (by over 25 percentage points) at a relatively small cost in specificity (less than 10 percentage points).

Acknowledgments

NCF is supported by a UK Medical Research Council Senior Clinical Fellowship. This work was undertaken at UCLH/UCL who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The DRC is an Alzheimer's Research UK Co-ordinating Centre and has received equipment funded by Alzheimer's Research UK.

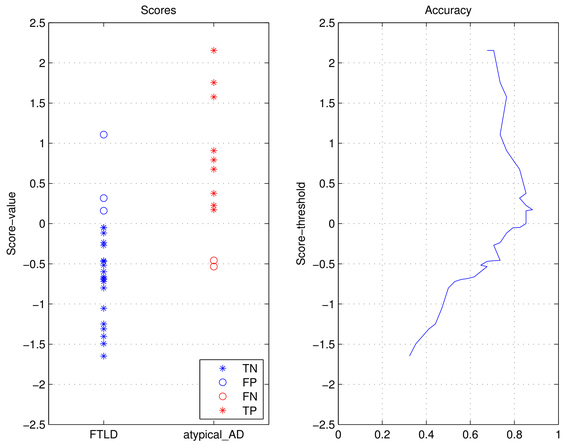


Figure 4: The modified SVM scores; accuracy with varying cut-point.

Conclusion

We have proposed a novel procedure to retrain the SVM threshold inside the CV loop, which can simultaneously achieve high sensitivity and specificity on unbalanced data, without compromising accuracy. Software will be made openly available as a toolbox for SPM: `http://www.fil.ion.ucl.ac.uk/spm/ext/#CLASS`

References

M. Lehmann *et al.*, "Reduced cortical thickness in the posterior cingulate gyrus is characteristic of both typical and atypical Alzheimer's disease." *J Alzheimers Dis*, vol. 20, no. 2, pp. 587–598, 2010. http://dx.doi.org/10.3233/JAD-2010-1401

Contact details

For poster-reprints or to enquire about the forth-coming software toolbox, please email `Gerard.Ridgway@ucl.ac.uk`