Comparing Cross Validated Classification Accuracies for Alternate Predictor Variable Weighting Algorithms

Mary G. Lieberman **John D. Morris**
Florida Atlantic University

The present research contrasts the effectiveness of four predictor variable weighting algorithms with respect to cross-validated accuracies in classification problems. Ordinary Least Squares Regression (OLS), Ridge Regression (RR), Principle Components (PC), and Logistic Regression (LR), are the techniques that were contrasted on 24 real data sets in terms of optimizing cross-validated classification accuracies. LR was best in only 1 data set, PC was best overall in 16%, RR was best in 8%, and OLS was best in 8% of the data sets.

This investigation contrasts four weighting algorithms for classifying subjects into a priori groups based upon classification accuracy. Ordinary Least Squares Regression (OLS), also the same as classification using a linear predictive discriminant analysis or in the case of two groups, Fisher's (LDF); Ridge Regression (RR); Principle Components (PC); and Logistic Regression (LR), are the techniques that were compared with respect to their cross-validated classification accuracies in real data.

In a regression context, Darlington (1978) posited that cross validation accuracy is a function of R^2, N, VC, where R^2 represents the squared multiple correlation, N is the sample size, and VC is defined as the validity concentration. In Darlington's formulation, validity concentration was used to describe a data condition in which the principal components of the predictors with large eigenvalues also have large correlations with the criterion. Thus, validity concentration requires some degree of collinearity. Darlington suggested that the most useful statistical techniques for practical prediction problems, as in personnel selection, may be ridge regression and Stein-type regression. These combine the sensitivity of multiple regression with the resistance to sampling error of other techniques—notably rational (clinical) weights and weights determined by simple correlations. Darlington stated that the new techniques are not recommended for theoretical modeling work because they yield biased estimates of the true least squares weights, typically have higher expected squared errors for estimating some weights, and do not allow the use of ordinary confidence bands or significance tests. Nevertheless, he recommended the use of ridge regression as best for most classification problems.

Morris (1982) re-examined the performance of ridge regression from a different methodological perspective using the same data structures on which Darlington (1978) demonstrated the technique's superiority. Contrary to Darlington's suggestions, Morris (1982) found that ridge regression was never the most accurate prediction technique, although least squares weights, as well as all of the other non-least-squares techniques, were most accurate in some data configurations.

Further, Morris (1983) examined Darlington's (1978) suggestion to utilize a "shrunken inter-correlation matrix" as the input to an ordinary stepwise regression program to accomplish a stepwise ridge regression solution. The algorithm that Darlington suggested calculates the portions of predictable criterion variance attributable to ridge weighted variable subsets incorrectly, causing inappropriate predictor variable subsets to be selected. An alternate stepwise ridge regression procedure is suggested by Morris (1983).

Through simulation, Morris (1982) showed that as R^2 decreases, N decreases, and the VC increases, Ridge Regression becomes better than Ordinary Least Squares but, as well, Reduced Rank, Equal Weighting, and other techniques become better than Ridge. In several studies, Morris (1982) and Morris and Huberty (1987) found that the performance of Ridge Regression was inferior to that of Ordinary Least Squares, Principal Components, Reduced Rank, and Equal Weighting in all but a few data structures.

In fact, there is some evidence that cross-validated R^2 becomes better with increased VC, even better than the R^2 of OLS at low VC. Because Validity Concentration requires collinearity, the interest might be in examining whether collinearity can, under some circumstances, be helpful to prediction. The present research seeks to contrast Logistic Regression, as a popular classification technique frequently proffered in the literature, with the prior three methods examined in Morris and Huberty (1987).

Method

A similar comparison (Morris & Huberty, 1987) examined only the OLS, RR, and PC methods. Logistic Regression will expand that coverage. Twenty four real data sets with varying degrees of group separation were analyzed using these four methods to ascertain differences in classification accuracies. All predictor weighting algorithms were cross validated using the Leave-One-Out technique. This algorithm is executed by alternately predicting each subject's group membership from the equation generated from the predictor and criterion scores of all other subjects. The resulting hit-rate over all subjects serves as a criterion for cross-validation accuracy.

Results

The present research seeks to expand upon prior work investigating the effects of three weighting algorithms on classification accuracies: OLS, Ridge, and PC. In those simulation studies, all methods performed better with increasing sample size, larger population multiple correlations, and large degrees of group separation. OLS performed better with smaller levels of validity concentration. As VC increased, the performance of Ridge Regression was superior, and, at very high levels of VC, Principal Components Regression was superior. It is salient to note that in larger samples, this trend was delayed (Morris & Huberty, 1987). Overall, non-OLS methods performed best, or with increased accuracy, in small samples. It should also be noted, however, that even at high levels of VC, and with significant differences in classification accuracies, the differences were often small (Morris & Huberty).

The finding with real data mirrored the simulation results, but with the focus on contrasting results for specific data sets; not a general contrast of methods. Table 1 reflects the results of the contrast in cross-validated hit rates for 24 real data sets with varying degrees of group separation, numbers of subjects, variables, and data co-variance matrices. As can be seen, LR, the additional method being contrasted, is not present in the first 3 data sets. Overall, LR is best in only 1 data set (i.e., # 15 Block 3 & 4). It is tied with other methods in 6 (29%) of the data sets. It is second best in 7 (29%) of the data sets, third in 2 (13%) of the data sets, and worst in 5 (21%) of the data sets. In two of these data sets (i.e., #6 Bisbey 1 & 2 and # 10 Rulon 1 & 3), LR performed the worst of all four of the methods. All methods performed equally well at a 79% hit rate in the #9 Demographics #2 data set. PC was best overall in 4 (16%) of the data sets and tied for best in 3 (13%) of the data sets. RR was best overall in 2 data sets (8%) and tied for best in 8 (33%) of the data sets. Finally, OLS was best overall in 2 (8%) and tied for best in 8 (33%) of the data sets.

Discussion

To summarize, the present research contrasted four predictor weighting algorithms: Ordinary Least Squares Regression, Ridge Regression, Principle Components, and Logistic Regression. The purpose of the study was to enhance researchers' methodological toolbox with the most accurate methods for selecting predictor variable weights in a cross-validated context. Subsequently, the weights chosen should yield greater classification accuracy for specific real data sets under investigation.

References

Darlington, R. B. (1978). Reduced variance regression. *Psychological Bulletin, 85*, 1238-1255.

Morris, J. D. (1982). Ridge regression and some alternative weighting techniques: A comment on Darlington. *Psychological Bulletin, 91*, 203-210.

Morris, J. D. (1983). Stepwise ridge regression: A computational clarification. *Psychological Bulletin, 94*, 363-366.

Morris, J. D., & Huberty, C. J. (1987). Selecting a two-group classification weighting algorithm. *Multivariate Behavioral Research, 22*, 211-232.

Send correspondence to: Mary G. Lieberman
Florida Atlantic University
Email: mlieberm@fau.edu

Table 1. Prediction Methods' PRESS Performance: Proportion of Hits

#	Data Set Source				Method			
		D	k	N/p	OLS	Ridge	PC	LR
1	Fisher 1 & 3	13.97	0.001	100/4	1.00	1.00	1.00	
2	Fisher 1 & 2	10.16	0.002	100/4	1.00	1.00	1.00	
3	Bisbey 1 & 3	5.12	0.033	72/13	0.97	0.97	0.92	
4	Fisher 2 & 3	3.77	0.012	100/4	0.97	0.97	0.81	0.97
5	Rulon 1 & 3	2,93	0.010	152/3	0.93	0.93	0.91	0.91
6	Bisbey 1 & 2	2.89	0.071	116/13	0.89	0.89	0.87	0.88
7	Bisbey 2 & 3	2.41	0.099	118/13	0.89	0.87	0.87	0.89
8	Talent 3 & 5	1.97	0.164	127/14	0.79	0.79	0.79	0.80
9	Demographics #2	1.88	0.034	279/8	0.79	0.79	0.79	0.79
10	Rulon 2 & 3	1.87	0.023	159/3	0.83	0.84	0.84	0.82
11	Rulon 1 & 2	1.74	0.022	179/3	0.81	0.81	0.80	0.80
12	Talent 1 & 5	1.72	0.116	177/14	0.75	0.75	0.72	0.73
13	Demographics #3	1.36	0.064	279/8	0.73	0.72	0.66	0.74
14	Talent 1 & 3	0.89	0.839	116/14	0.62	0.70	0.70	0.62
15	Block 3 & 4	0.85	0.307	76/4	0.67	0.67	0.58	0.69
16	Block 1 & 2	0.84	0.308	77/4	0.66	0.67	0.69	0.66
17	Block 1 & 4	0.81	0.325	78/4	0.58	0.59	0.55	0.58
18	Block 1 & 3	0.74	0.387	78/4	0.62	0.60	0.58	0.60
19	Warncke 1 & 3	0.69	0.950	105/10	0.61	0.58	0.59	0.61
20	Block 2 & 3	0.64	0.550	75/4	0.55	0.55	0.59	0.55
21	Block 2 & 4	0.52	0.814	75/4	0.59	0.59	0.64	0.59
22	Demographics #1	0.50	0.477	279/8	0.59	0.58	0.57	0.58
23	Warncke 1 & 2	0.48	1.749	112/10	0.47	0.54	0.58	0.48
24	Warncke 2 & 3	0.45	2.635	87/10	0.41	0.46	0.43	0.42