

Problem Set 10 - ECON 5253

Thomas Mondry

April 18, 2022

1 Table of results

Algorithm	Accuracy	Standard error	Hyperparameters
Penalized logit (LASSO)	0.85260	0.00204	$\lambda = 1\text{E-}10$
Decision trees	0.86842	0.00053	minimum split size = 10, maximum depth = 15, cost complexity = 0.01
Neural network	0.85675	0.00074	$\lambda = 1$, 4 hidden units
k-nearest neighbors	0.84339	0.00179	$k = 30$
Support vector machine	0.86396	0.00062	cost = 1, $\sigma = 0.25$

2 Commentary

As shown in the table, most of the ML algorithms achieved similar accuracy levels on the test data. k-nearest neighbors is probably the least efficient approach to this problem – it has a nominally lower accuracy score than all of the other algorithms, and it evaluates 30 neighbors in doing so, which is the maximum number we allowed it to try. SVM is slower to train than any of the other models by at least an order of magnitude, and while it is a strong model, the decision tree model works marginally better & is much less computationally intensive. I would go with the decision tree based on its performance here, although if time or computational resources were a concern, the penalized logit model works nearly as well and trains very quickly.