

Problem Set 10

Will Miller

April 2025

1 Model Evaluation

alg	.estimate	penalty	hidden_units	tree_depth	min_n	cost_complexity	neighbors
logit	0.85	0.00					
tree	0.87			15.00	10.00	0.00	
nnet	0.83	1.00	1.00				
knn	0.84						30.00

Table 1: Model estimates and hyperparameters

Overall, my tree model was the most accurate despite all models being pretty close in terms of accuracy. This isn't super surprising to me when considering the features of tree models and our lack of hyperparameter tuning in our preprocessing stage. Because they are nonparametric, tree models are able to incorporate nonlinearities and interactions between variables without having them explicitly specified. This property makes them prone to overfitting, but by regularizing the trees using cross validated hyperparameter tuning, we take steps to alleviate that concern. For that reason, it is unsurprising

I am also not shocked that kNN performed the worst here. We did not scale our variables, meaning that kNN—which uses distance to make its predictions would be particularly sensitive to this choice.

I was unfortunately unable to SVM to run on my computer. I believe this is a result of the computational complexity of the algorithm.