PS10-Onishi

Saryu Onishi

April 2023

1 Out-of-sample performance comparison

The .estimate column represents out-of-sample performance. Based on the results shown in the table, it can be concluded that the decision tree algorithm may be the best-performing algorithm. The second best is the logistic regression algorithm, as it maintains a good out-of-sample performance without losing complexity.
The next nearest neighbour algorithm is the worst performing, relatively. However, all four models from the four algorithms have produced models with relatively similar out-of-sample performance.
The SVM model was omitted because I could not resolve an error interrupting the tuning process.

penalty	.estimate	alg	cost_complexity	tree_depth	min_n	hidden_units	neighbors
0.00	0.85	logit					
	0.87	tree	0.00	10.00	50.00		
1.00	0.85	nnet				1.00	
	0.84	knn					28.00