

Problem Set Ten

Devin Williams

April 15, 2025

1 Question 9

Table 1: Optimal Tuning Parameters and Out-of-Sample Performance

Algorithm	Accuracy	Parameters
Decision Tree	86.8%	cost_complexity = 0, tree_depth = 20, min_n = 10
SVM	86.4%	cost = 1, rbf_sigma = 0.25
Logistic Regression	85.3%	penalty = 0
KNN	84.3%	neighbors = 30
Neural Network	82.9%	penalty = 1, hidden_units = 5

- How does each algorithm's out-of-sample performance compare with each of the other algorithms?
 - Decision Tree
 - ◊ The Decision Tree at 86.8 percent achieved the highest accuracy. This suggests that the relationship between the predictors and high-income status involves non-linear interactions. The parameters listed show that a deep tree with minimal pruning was the most effective.
 - SVM
 - ◊ SVM performed nearly as well as the decision tree, with less than 1 point of difference. This shows that the data has a clear separation that can be modeled using kernel methods.
 - Logistic Regression
 - ◊ Despite its overall simplicity, logistic regression also performed quite well. This suggests that linear decision boundaries are still reasonably effective for this task. This also shows that some of the relationships between predictors and the outcome may be almost linear in nature.

- KNN
 - ◊ This algorithm with optimal 30 neighbors performed the second worst, although not by much. The relatively high value of K suggests that local noise in the data benefits from substantial smoothing.
- Neural Network
 - ◊ The neural network with 5 hidden units showed the lowest performance, though still had a good result overall. This might be due to the limited size of the network.
 - ◊ Overall, the performance of all 5 are very small (with a 4 percentage point difference from best to worst). This shows that all 5 provide reasonable predictive power for this task. For this problem, the decision tree shows the best performance, suggesting that explicit interpretable decision rules may be particularly suited to income prediction based on demographic and employment characteristics.