# PS10_Zhang

Haotian Zhang

April 2025

## 1 Answer for Question 9

| Algorithm | Best Tuning Parameters | CV Accuracy | Test Accuracy |
|---|---|---|---|
| Logistic Regression | penalty = $1 \times 10^{-10}$ | 0.8457 | 0.8526 |
| Tree Model | cost_complexity = 0.001, tree_depth = 20 | 0.8607 | 0.8690 |
| Neural Network | penalty = 1 | 0.8476 | 0.8539 |
| kNN | neighbors = 28 | 0.8368 | 0.8437 |
| SVM (RBF Kernel) | cost, rbf_sigma = 1 | 0.8530 | 0.8641 |

Table 1: Optimal tuning parameters and accuracy for each model

The out-of-sample performance of the five classification models varies, with the Tree Model algorithm achieving the highest accuracy at 0.8690. This is closely followed by SVM with an accuracy of 0.8641, and the Neural Network at 0.8539. The Logistic Regression model performs comparably with 0.8526, while kNN has the lowest accuracy at 0.8437.

This suggests that models capable of capturing non-linear interactions (like Tree Model and SVM) may better handle the structure in the income dataset than linear models like logistic regression or distance-based methods like kNN.