

ECS 111 Homework Assignment #2

Kevin Oghalai

May 7, 2025

1 Introduction

The data was loaded in and then analyzed to determine possible issues that might arise from missing values and the various types of data. In this case, there were multiple missing values in addition to many categorical variables, which could not be directly used for analysis. To fix this, one-hot encoding was used. Although data could have been filled in with the mean or median of the column, this would not have worked for categorical variables, and in the end, this did not end up being necessary.

2 Data Processing

A logistic regression model was used to predict the probability of a person having an income of over \$50,000. This algorithm comes with options to regularize the data, in addition to loss functions, optimizers,