CS235 Fall'23 Project Proposal: Use Deep Learning to Predict Car Sales Price

LIAM Y. HSIEH (862395843), University of California, Riverside, U.S.A

This project focuses on predicting car sales prices using Deep Learning and compares them to a baseline linear regression model. The dataset, obtained from Kaggle, comprises 500 records with features like customer demographics and financial information. The ANN model is structured with an input layer, two hidden layers using ReLU activation, and an output layer for regression. Evaluation employs Mean Absolute Error (MAE) as the primary metric. Liam Hsieh is the sole contributor for this project, aiming to optimize predictive accuracy and provide valuable insights for the automotive industry.

Additional Key Words and Phrases: Car sales price prediction, Deep Learning, Neural Networks, Mean Absolute Error (MAE)

1 INTRODUCTION

The proposed project aims to leverage Deep Learning to predict car sales prices based on a dataset obtained from Kaggle. The dataset consists of 500 records with 9 columns, encompassing customer details such as name, email, country, gender, age, annual salary, credit card debt, net worth, and car purchase amount. The objective is to explore the potential of ANN models in predicting car sales prices compared to a baseline model using linear regression. Linear regression will serve as the benchmark model, and various ANN structures will be evaluated to identify the most suitable and accurate predictive model. By harnessing the capabilities of ANN, we aim to enhance the accuracy and precision of car sales price predictions, providing valuable insights for the automotive industry and aiding customers in making informed purchasing decisions.

1.1 Project Type

The proposed approach falls under the category of a Software project as it implements existing methodologies and compares the results obtained.

2 PROBLEM DEFINITION

Accurately predicting car sales prices is a critical task that significantly influences purchasing decisions for both consumers and dealerships. The existing methods often rely on traditional statistical approaches like linear regression, which may not fully capture the complexities and non-linearity present in the data. The problem at hand involves utilizing Artificial Neural Networks (ANNs), a powerful tool in machine learning, to predict car sales prices. ANN models can potentially grasp intricate patterns and relationships within the dataset, enabling superior predictive performance compared to conventional linear regression methods. By exploring different ANN architectures and comparing them with a baseline linear regression model, we aim to identify the most effective approach for predicting car sales prices accurately.

2.1 Dataset

The dataset utilized in this project is sourced from Kaggle [1] and is centered around predicting car sales prices using machine learning techniques. The dataset contains 500 records with 9 columns, featuring both predictors and the target variable.

Predictors:

[1] https://www.kaggle.com/datasets/yashpaloswal/ann-car-sales-price-prediction

- Customer Name: The name of the customer.
- Customer Email: The email address of the customer.
- Country: The country of the customer.
- Gender: Gender of the customer (binary encoding: 0 for male, 1 for female).
- Age: Age of the customer.
- Annual Salary: The customer's annual salary.
- Credit Card Debt: The amount of credit card debt the customer has.
- Net Worth: The net worth of the customer.

Target/Label:

- Car Purchase Amount: The amount a customer spends on purchasing a car.

The dataset provides a diverse range of customer information that can be utilized to train and evaluate machine learning models for predicting car purchase amounts. The inclusion of both demographic and financial features allows for a comprehensive analysis, which is crucial in building accurate prediction models for car sales prices.

3 PROPOSED APPROACH

The proposed approach involves initially splitting the dataset into a training set (80%) and a testing set (20%). The core of the project revolves around leveraging Artificial Neural Networks (ANNs) to predict car sales prices. The ANN architecture comprises an input layer representing selected features, two hidden layers activated using Rectified Linear Unit (ReLU) functions, and an output layer for regression, employing the mean squared error loss. The Keras library, built on TensorFlow, will be utilized to implement the ANN model. Specifically, the output layer will employ a linear activation function to predict numeric car purchase amounts. Different activation functions will be explored and compared to determine the most effective model configuration. The ANN model will be compiled using a defined optimizer and loss function, and subsequently trained on the training set. Evaluation of the ANN's performance will be conducted using the testing set. Additionally, a baseline linear regression model will be implemented and compared with the ANN model to assess and compare predictive performances, ultimately selecting the most accurate and efficient model for car sales price prediction.

4 EVALUATION METRICS

For assessing the accuracy of our car sales price prediction models, Mean Absolute Error (MAE) will be the primary evaluation metric. MAE calculates the average absolute differences between the predicted and actual car purchase amounts. A lower MAE signifies a more accurate prediction model, where smaller absolute differences indicate closer alignment between predicted and observed prices.