

Prediction of a car's selling price

Introduction

(Please don't grade the introduction, it's there to give a little bit more information overall and just a draft)

When selling a quite old and used car privately, and without knowing too much about cars overall, it can be quite hard to set a good price for the car. The aim of this project is to estimate my car's price based on the prices of similar cars as mine that are for sale at two different Finnish car sale sites.

A good price for an older and used car is a price where the seller does not make too much loss of the car's present value, but the price should be low enough to attract potential buyers.

Problem formulation

The data points of this machine learning problem are cars. The prices and information of the cars were collected by hand from Nettiauto [1] and Autotalli [2], that are Finnish car selling sites, during 23-24.1.2022. Only cars that are sold by car selling companies were taken into account when data was collected, since they have much more and more accurate knowledge of prices that cars should be sold for, compared to when selling privately.

When thinking about features for cars, different cars that are of the same model can have different devices and a different amount of them e.g., for safety, comfort, and electronical devices that affect the price of the car positively even if the car is older and more driven and vice versa. In other words, this machine learning problem could be quite complex if all devices of the cars were taken into account. Therefore only a few important features were chosen for this project. As features of the data points were chosen the driven distance (km), age of car (years), fuel (gasoline or diesel) and engine volume (cm³). As the label was chosen the price of a car (€). The data set contained 416 data points, and no information was missing regarding the features and labels.

These features were chosen, because the driven distance and age of the car affects the price most, even if the car has many fancy extra devices. The engine volume was added as a feature since there exists similar cars as the one that's price is going to be predicted, but with smaller or larger engines which affects the price of a car. The car which price is predicted is diesel-powered, but both gasoline and diesel-powered cars were included to the data set, because enough diesel-powered cars were not for sale to make a big enough data set. All of the cars in the data set have a manual gearbox, as the one which price is predicted.

The aim of this project is to predict the price of a car, based on the collected data set of similar cars, by using linear regression because it is a simpler method than many other machine learning methods. Due to multiple features, the simplicity of linear regression makes it easier to predict the car price, at least for a beginner in machine learning. Based on common knowledge of cars, it can be said that the more driven and older the car is, the lower the price of it is, but a larger engine raises the price of the car. It is hard to say though how the price is affected depending, whether the car is gasoline or diesel-powered. It will be interesting to see how an individual feature of a car affects its price and it can be studied through the coefficients of the linear regression model.

References

1. <https://www.nettiauto.com/>
2. <https://www.autotalli.com/>