Cars prediction price

Using machine learning linear regression models

Presented by : Omnia Abduo Mostafa Eid

Year :2023

Supervisor: Dr / Abeer El-desouky

Programme: pre master of computer science

Abstract

Cars prediction price based on set of feature such as car engine size, car height, car length and car width , the more this sets them apart from comparable cars, the harder they become to evaluate with traditional methods. Using machine learning algorithms (boosting model) to better utilize data on all less common feature of a car can more accurately assess the value of a vehicle. this study compares results after using linear regression model and after using boosting algorithm. The results show that boosting algorithms improve the accuracy of data training and data testing.

Keywords: machine learning ,car prediction price, linear regression, data train ,data test and boosting algorithm .

---

(1) Introduction

1.1 background

the used car prediction price based on its features . the better features , the more price. There are charts show relationship between car price and every features individually .

1.2 overall aim

the purpose of this project is to evaluate several different machine learning models for used car price prediction and draw conclusions about how they behave.

1.3 problem statement

there are several machine learning regression models that can be applied to price prediction. This work will investigate which one offers the best performance according to several criteria. I tried to solve this problem by studying each model individually and finally used the model gives me the highest accuracy on data that I used (linear regression).

1.4 Research Questions

(1) Which ML model and parameters gives the best overall accuracy in making price predictions for used cars?

(2) Which ML model can most accurately assess the depreciation of car over time?

Question 1 will determine which several algorithms gives the best performance . Question 2 will examine and compare the behavior of the algorithms to suggest which can best assess depreciation over time

---

(2) Material

2.1 Machine learning

It is a subfield of artificial intelligence that works with algorithms and technologies to make useful inferences from data . ML algorithms are suited to

problems containing large amounts of data which not be possible to process without such algorithms.

2.2 Linear regression

It is a technique to estimate the linear relationship between each of number of independent variables and a dependent variables.

2.3 Boosting algorithm

It is a method used in machine learning to reduce errors in predictive data analysis.

---

(3)Implementation /Design

3.1 Dataset

Data consist of 99 raw and 5 columns

	enginesize	carlength	carwidth	carheight	price
0	130	168.8	64.1	48.8	13495
1	130	168.8	64.1	48.8	16500
2	152	171.2	65.5	52.4	16500
3	109	176.6	66.2	54.3	13950
4	136	176.6	66.4	54.3	17450

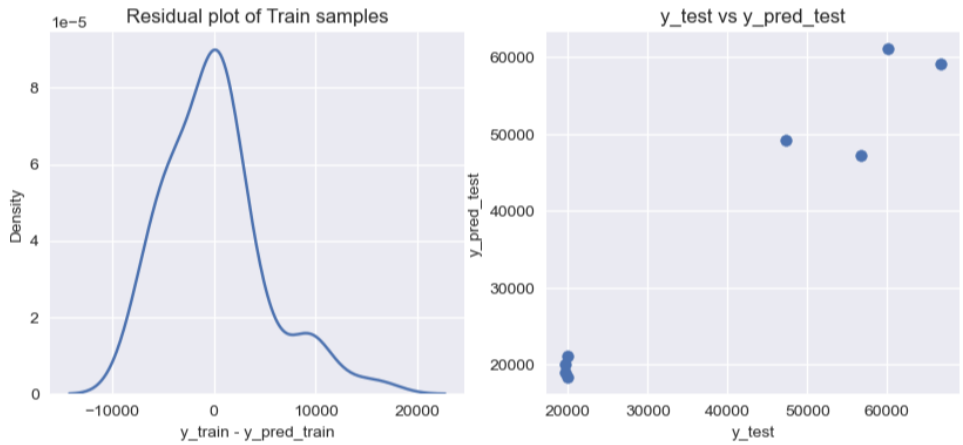
3.2 model creation evaluation

* Linear regression model

Using to show the relationship between independent and dependent variables

Results

```
Train R2-score:  0.91
Test R2-score:  0.95
Train CV score:  [0.89008164 0.9260122  0.8557138  0.93266952 0.75807153]
Train CV mean:  0.87
```

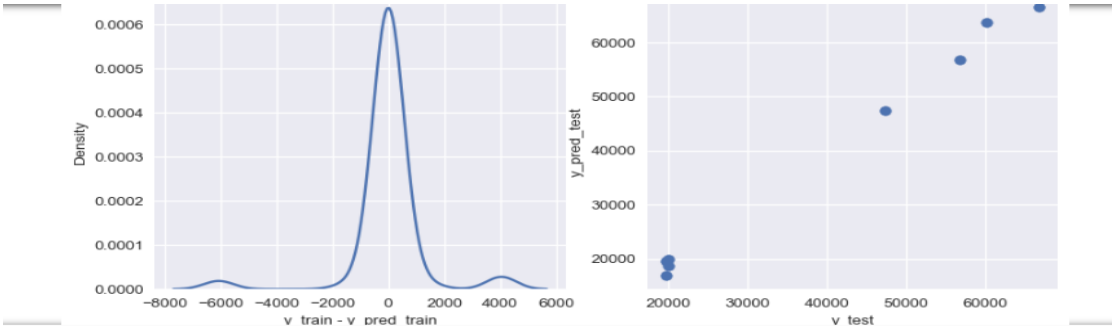


*Gradient boosting

Using to improve the accuracy

Results

```
Train R2-score:  0.99
Test R2-score:  0.99
Train CV score:  [0.98307736 0.99517648 0.9088012  0.98358293 0.79234892]
Train CV mean:  0.93
```



---

(4)Conclusion

This project will summarize the work as a whole by conclusively answering the research questions .

Question 1 will determine which several algorithms gives the best performance . Question 2 will examine and compare the behavior of the algorithms to suggest which can best assess depreciation over time.