

Report

I spent a large amount of time trying to set up everything to work properly, however, I have to admit that I failed when trying to test the prediction performance using script because of the security issue related to API key and OAuth 2.0. To figure out how well Google Prediction API worked, I have to use the way provided in Google's document other than using Python.

In my experiment, I found the data sheet with more than 800 entries which describes the relationship between the prices of the cars against the properties of the associated cars such as number of doors, mileage, car style and so on.

When the data is imported into Google Prediction API, API selected out the correct training model which is regression one, however, it failed to give out the satisfactory result when trying to predict the price of a car with specific property values that I assigned.

As mentioned before, there are 10 properties assigned to the cars which can have impact, either positive or negative on the prices of the cars. API worked perfectly to figure out which model should be implemented without my indication. However, it did not find out the correct weights that should be assigned to the properties when implementing the prediction task. Here I assigned the a group of values to it and require the prediction result. I set mileage is 30000, make is "Buick" with model "Century" and trim "Sedan 4D", its type is set as "Sedan", the numbers for cylinder, liter, doors are 6,3,1,4 respectively. For the logically value associated with cruise, sound and leather are assigned as 1,1,1 which indicate true. From the resulted price predicted with the trained model, we can see that the car is worth 19674.403780 which is much higher than the exact same car with lower mileage. Intuitively, this is not correct. Similar results can be found when other values are implemented.

In my opinion, there are several reasons to cause this. Firstly, there are too many properties that should be implemented into the model, which made the model too complicated. Such a model might be overfitting which means that it may work perfectly on the training data, but failed to work in prediction. Secondly, suppose all of the properties should be taken into consideration, the model might not have enough information on the importance of different properties, which results in the violation to our intuitions.

Taking the potential problems mentioned above into consideration, I would personally firstly review the training data to make sure it is logically correct and can provide me enough information. For example, a very important property which is the year of the car is omitted in this data. Secondly, the problems can be divided into smaller ones, or at least the number of properties can be reduced using regularization method.

Request

```
POST https://www.googleapis.com/prediction/v1.6/projects/977260187266/trainedmodels/carprice/predict?key={YOUR_API_KEY}

{
  "input": {
    "csvInstance": [
      "30000,\"Buick\", \"Century\", \"Sedan 4D\", \"Sedan\", 6, 3, 1, 4, 1, 1, 1"
    ]
  }
}
```

Response

```
200 OK
- Show headers -

{
  "kind": "prediction#output",
  "id": "carprice",
  "selfLink": "https://www.googleapis.com/prediction/v1.6/projects/977260187266/trainedmodels/carprice/predict",
  "outputValue": "19674.403780"
}
```