

Linear and Multiple Regression using scikit-learn

Arbaaz Khan

PhD Student, Computational EM Lab

Dept. of Electrical & Computer Engineering

McGill University

May 13th, 2019

Outline

- Introduction
- Model Development
- Linear Regression
 - Implementation with scikit-learn
- Multiple Regression
 - Implementation with scikit-learn
- Conclusion

Introduction – Machine Learning

- Although Machine Learning has gained a lot of importance in recent times, it has been in existence since 1940.
- Some factors that are directly responsible for the growth in past few years are:
 - Computation Power (GPUs, ASICs, TPU).
 - Data.
 - Improvement in algorithms (CNN , LSTM).
 - Infrastructure & Frameworks (Git, scikit-learn, Tensorflow).



Siri



Google Assistant



Hey Cortana



Why should you study ML?

Neural nets have gained immense popularity in recent time for tasks such as Pattern Recognition, Image classification, Text summarization , etc.

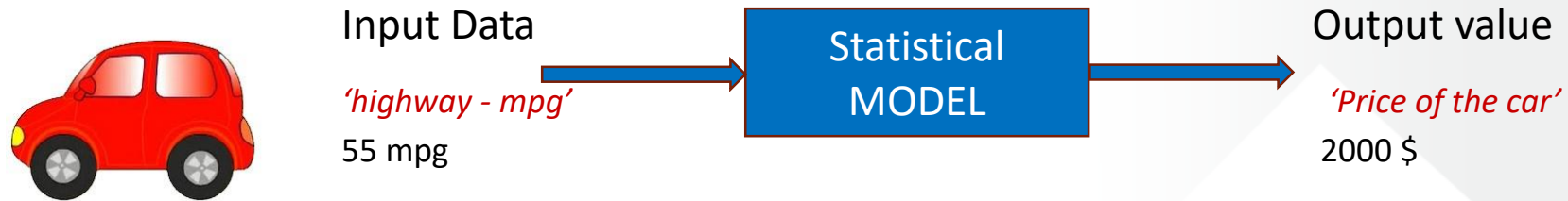
Why should we study Linear & Multiple regression?

Building blocks of neural nets.

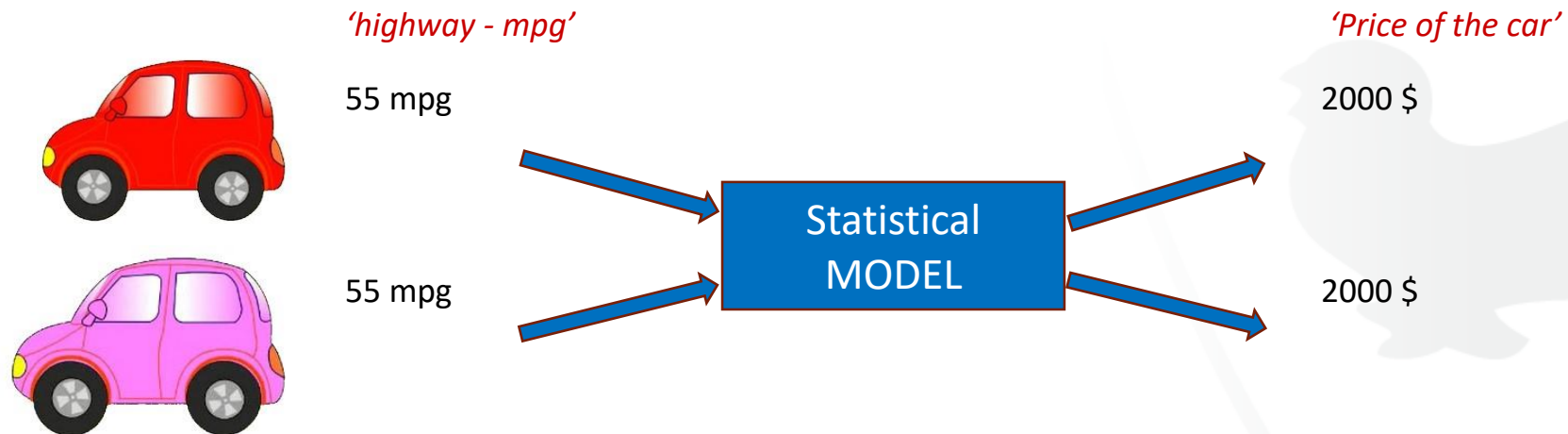
The Netflix logo, featuring the word "NETFLIX" in white, bold, sans-serif capital letters on a red rectangular background.

Model Development

- Predict the value of the car, given input data.



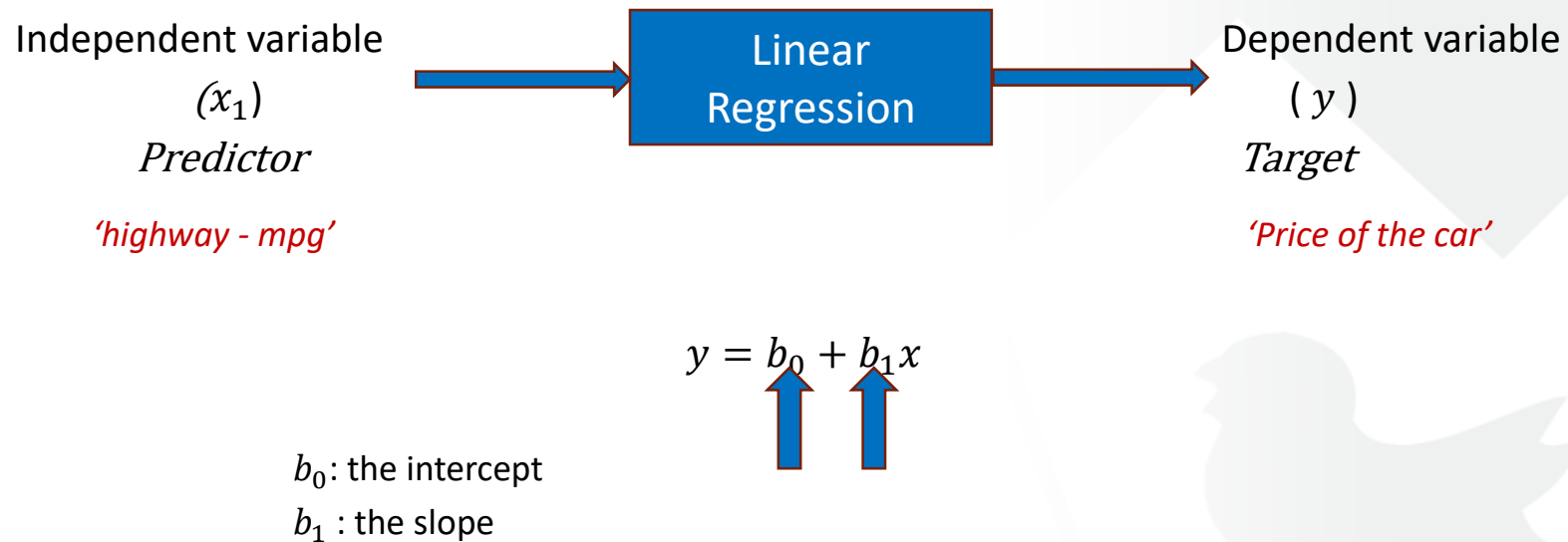
- More the relevant information, the better the accuracy.



Additional information such as the *'color of the car'* will be beneficial.

Linear Regression

Linear Regression uses one independent variable to make a prediction.



Linear Regression using scikit-learn



1. Import the *linear_model* from *scikit-learn*

```
from sklearn.linear_model import LinearRegression
```

2. Create a *Linear Regression* object using the constructor

```
model = LinearRegression()
```

3. Data is represented in machine readable format.

Highway – mpg (x)	Price of the car (y)
3	37486
7	35258
13	31230
22	22698

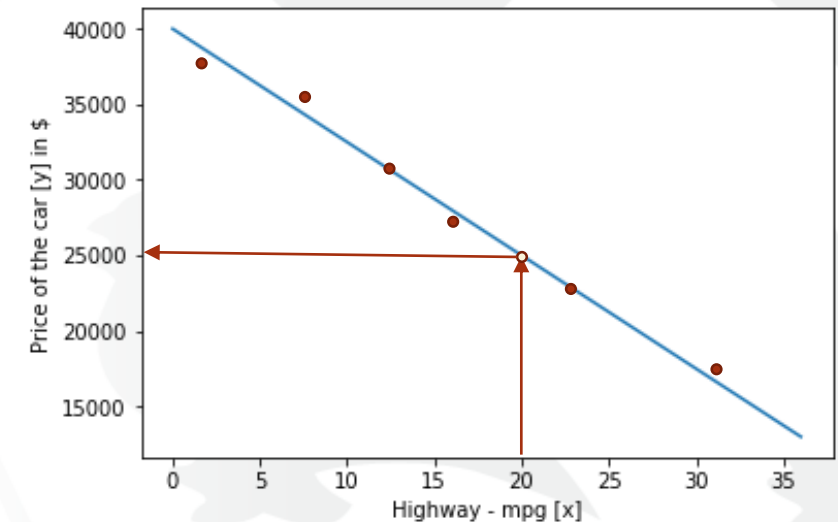
4. Use the method *model.fit()* to train the model.

```
model.fit(X, y)
```

$$y(x) = 40000 + 750x$$

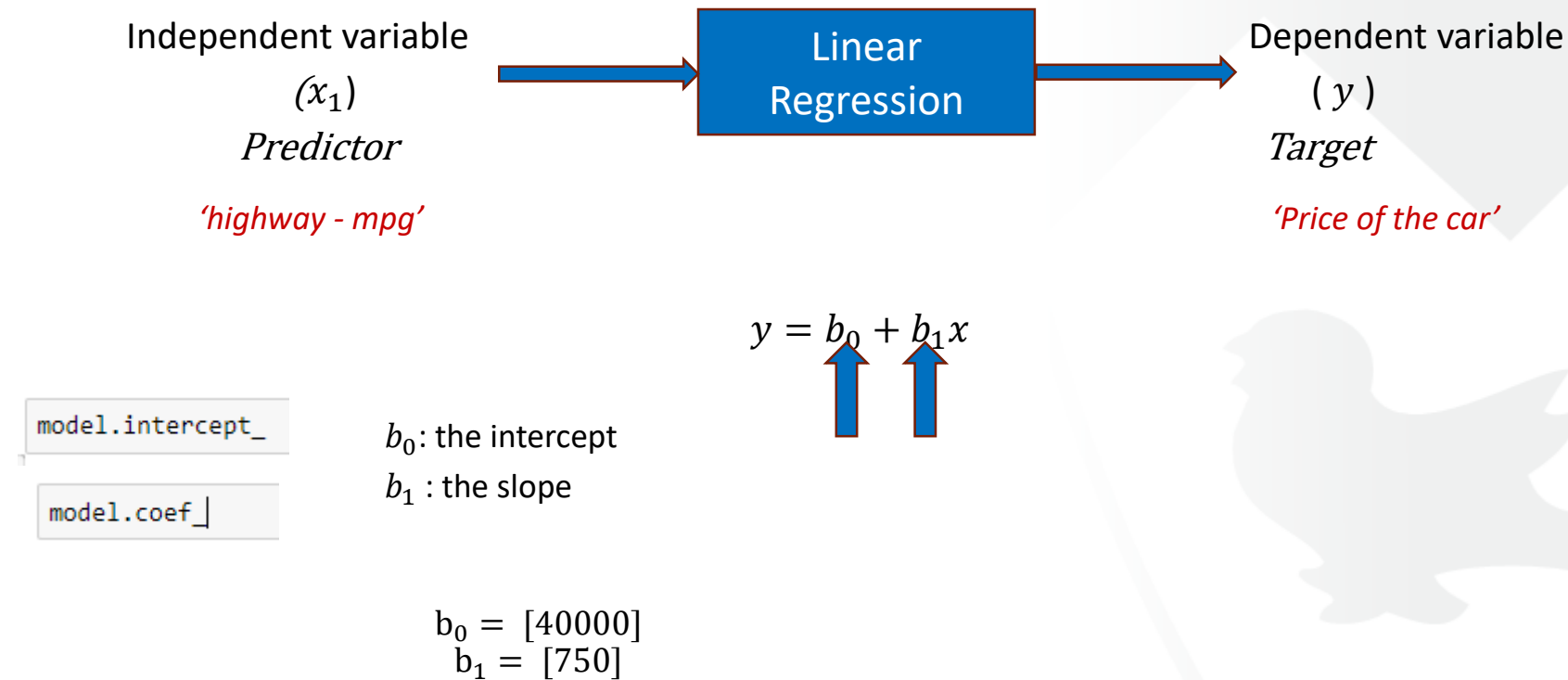
$$y(20) = 40000 + 750 * 20 = 25000 \$$$

```
Y-pred = model.predict(20)
```



Linear Regression - revisited

Linear Regression uses one independent variable to make a prediction.



Multiple Linear Regression (MLR)



- Multiple Linear Regression is used to explain the relationship between:

- A continuous variable (y)
- Two or more predictor variables(X)

- For 3 predictor variables: $y(X) = b_0 + b_1x_1 + b_2x_2 + b_3x_3$

where:

b_0 : intercept (X=0)

b_1 : coefficient of x_1

b_2 : coefficient of x_2 and so on....

- Train the model as before:

```
model.fit(X, y)
```

Remember that the predictor (X) here is a multi-dimensional vector

Highway – mpg (x_1)	Length (x_2)	Engine size (x_3)	Horse Power (x_4)	Price of the car (y)
3	540	740	150	37486

Linear and Multiple Regression using scikit-learn

Thank you!

Questions or Comments?

McGill University

Email address: arbaaz.khan@mail.mcgill.ca