Demand amount prediction for a retail product utilizing logistic regression

By
O.G.V.V Bandara

2023

Acknowledgement

This project was conducted as a passion project in order to understand the concepts of machine learning. Also to understand the logisitic regression model and it's applications.

Abstract

A dataset was made using excel such that the data corresponds to the data of a product in a country with rising inflation. The code was developed in python. The machine learning logistic regression model was utilized for the algorithm.

Contents

Chapter 1

Introduction

1.1 Background of study

1.1.1 Logistic regression

logistic regression models are models that have a certain fixed number of parameters that depend on the number of input features, and they output categorical prediction. Logistic Regression models are classification models; specifically binary classification models (they can only be used to distinguish between 2 different categories. the two categories can be plotted in the maxima and minima of the logistic function. The function with values between 0 and 1 can be used for probabilities. the 0 can be taken as less likely and 1 can be taken as more likely. the logistic functions where the values fall between -1 and 1 can be taken for binary classification where the +1 and -1 can be considered as the highes likely for a certain two types of binary categories.

1.1.2 Sigmoid function

The sigmoid function is a mathematical function that has an S shaped curve.

- logistic function
- hyperbolic tangent
- arctangent

All sigmoid functions have the property that they map the entire number line into a small range such as between 0 and 1, or -1 and 1 There are many different types of sigmoid functions, but they share some common similarities:

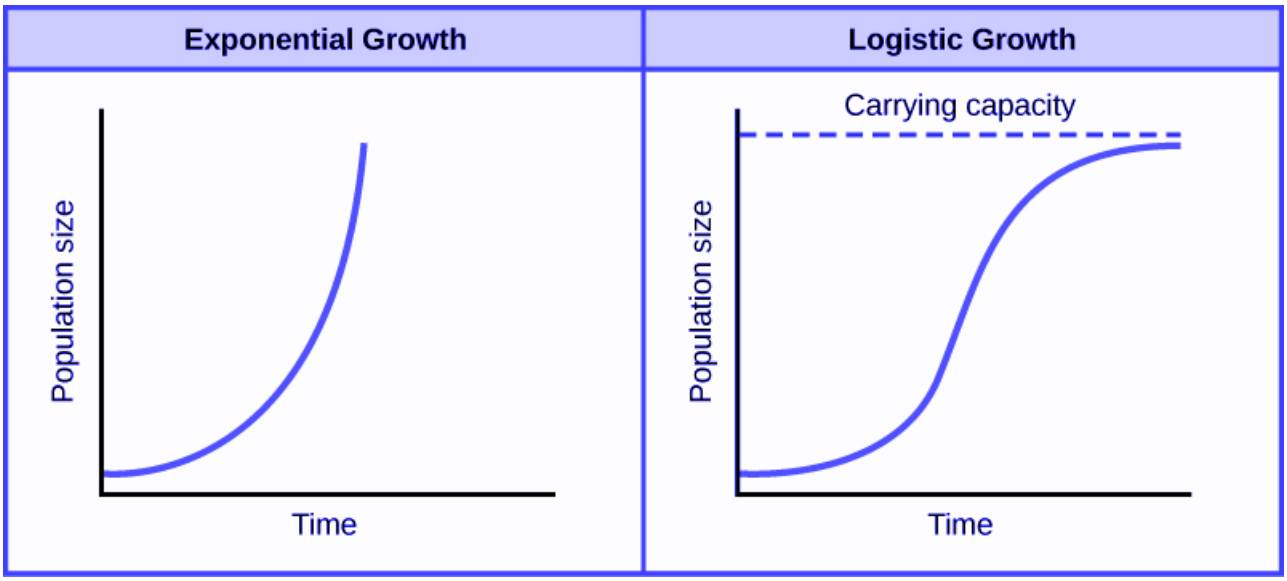
- Non-linear: Sigmoid functions are non-linear functions, meaning that the output is not proportional to the input. This allows them to model complex relationships between variables.
- Asymptotic: Sigmoid functions approach a horizontal asymptote as the input approaches infinity or negative infinity. This means that the output approaches a maximum value, which is often 1.

- Monotonic: Sigmoid functions are monotonic, meaning that they either increase or decrease monotonically as the input increases. This property makes them useful for modeling phenomena that have a threshold or tipping point.
- Continuous: Sigmoid functions are continuous functions, meaning that they have no abrupt jumps or discontinuities in their output.

Logistic function

The logistic sigmoid function or the function is also commonly referred to as the sigmoid function in machine learning,

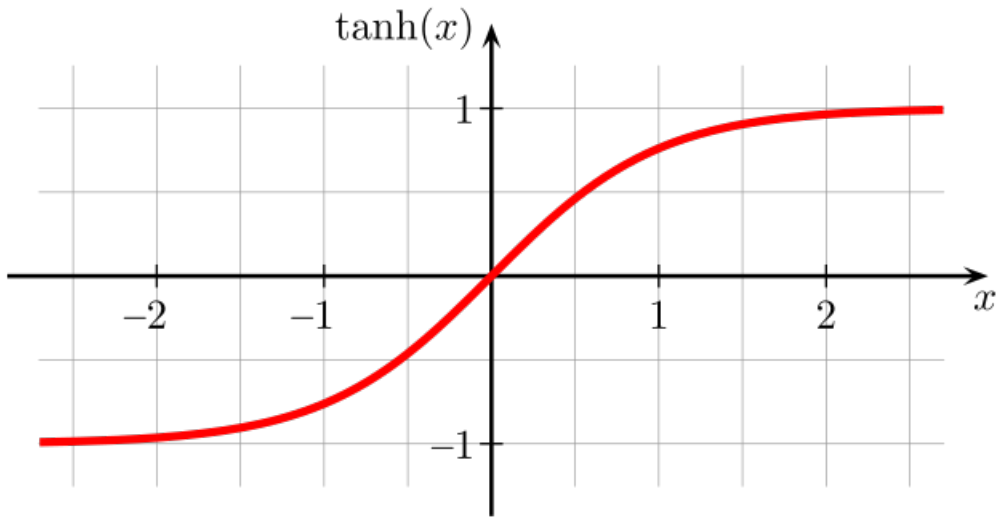
$$s(x) = \frac{1}{1 + exp(-x)} \tag{1.1}$$



hyperbolic tangent

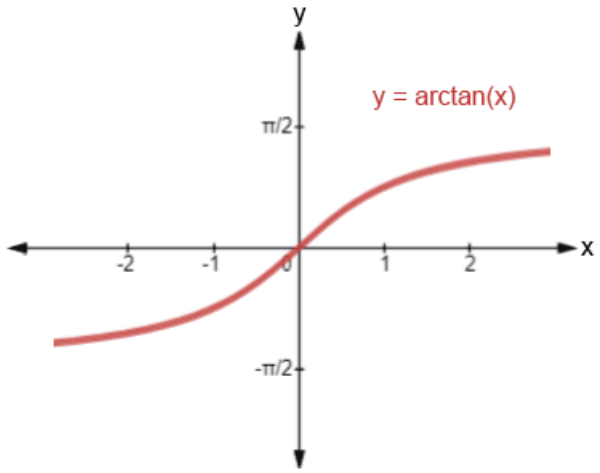
Another common sigmoid function is the hyperbolic function. This maps any real-valued input to the range between -1 and 1.

$$f(x) = \frac{exp(x) - exp(-x)}{exp(x) + exp(-x)} \tag{1.2}$$



arctangent

$$f(x) = arctan(x) \tag{1.3}$$



1.2 Objectives

- Test the ability of the logistic regression model to predict a value based on factors that change over time
- develop a code in order to predict the Demand price of future values accurately
- learn the functionalities of the scikit learn library in python

Chapter 2

Methodology

In this project the logistic regression machine learning algorithm is utilized to use past data in order to predict the demand of a product. The percentage of probability can be used to estimate whether the product will be highly demandable or less likely demandable.

- the data set was developed in the form of a csv file such that the data varies with time in a realistic nature of a country that has rising inflation and economic instability
- the code was developed in python using machine learning algorithms
- the function used here is the "logistic regression" fuction

Chapter 3

Conclusion

The logistic regression function is useful for binary classification and multiclass classification. Mostly the dependent variable should consist of integers or categorical data. The logisitic regression function is less accurate in predicting values that vary with time and depend on many features that vary with time.

Chapter 4

Appendix

```
from sklearn.linear_model import LogisticRegression
import pandas as pd

# load data from a CSV file
data = pd.read_csv('pricepred5.csv')
print(f"the imported dataframe is\n\n\n{data}")


#obtaining top 20 rows of csv
subset = data.loc[0:19, ['month', 'inflation','production price','Demand']]
print("\n\n\n")
print(f"the top 10 rows of the dataframe are:\n\n\n{subset}")


# splitting the data into features (X) and labels (y)
X = subset[['inflation', 'production price']] # select columns for features
y = subset['Demand'] # select column for labels

# creating a logistic regression model and fit it to the data
clf = LogisticRegression(max_iter=3000) #clf is now the model name
clf.fit(X, y)

# making predictions on new data
new_data = pd.read_csv('new_data2.csv')
print("\n\n\n")
print(f"the data set where the demand to be predicted is:\n\n\n{new_data}")
X_new = new_data[['inflation', 'production price']]
y_pred = clf.predict(X_new)

#printing the predicted values
print(y_pred)

# concatenate the array and dataframe along axis 1 (columns)
combined = pd.concat([new_data, pd.DataFrame(y_pred.T)], axis=1)
```

```
print(combined)

#the values to be predicted
y_new=data.loc[20:39, ['month', 'inflation','production price','Demand']]
print(f"the part of the data set that was predicted:\n\n{y_new}")

#seperating the column of the values to be predicted
y_topredict=y_new["Demand"]
print(f"the column to be predicted is:\n\n\n{y_topredict}")

#confusion matrix
from sklearn.metrics import confusion_matrix
confusion_matrix(y_topredict,y_pred)
```