S. B. JAIN INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NAGPUR.

Practical No. 3

Aim: Demonstrate the Problem related to Logistics Regression in Data Analytics.

Name of Student:

Roll No.: CS21

Semester/Year: VI / III

Academic Session: 2023 - 24

Date of Performance:

Date of Submission:

AIM: Demonstrate the Problem related to Logistics Regression in Data Analytics.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- Understand the use of odds, odds ratios and transformations in logistic regression.
- Logistic regression is a statistical analysis method to predict a binary outcome
- To measure the relationship between a categorical dependent variable and one or more independent variables (usually continuous) by plotting the dependent variables' probability scores.
- Able to calculate both simple and multiple regression models. You will learn how to assess the model's "fit", test model assumptions, and transform predictor and response variables to improve outcomes

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement: Computer System

Software Requirement: Google Colab, Kaggle Dataset

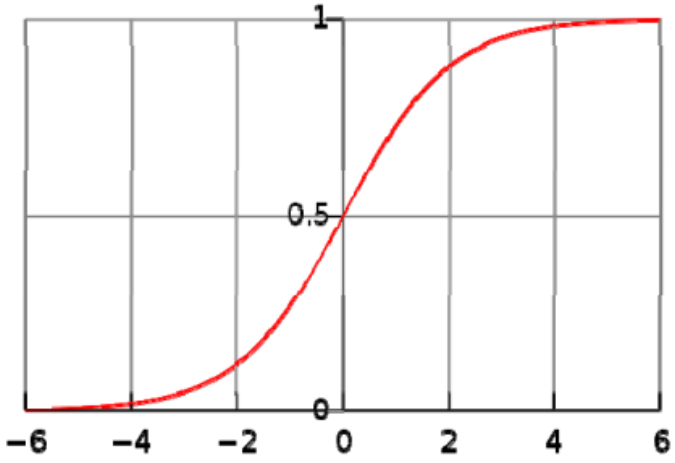
THEORY:

Logistic regression

- Name is somewhat misleading. Really a technique for classification, not regression. technique for classification, not regression.
- "Regression" comes from fact that we fit a linear model to the feature space
- Involves a more probabilistic view of classification

$$p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \qquad \text{logistic function}$$

Standard logistic function



Using a logistic regression model

Can interpret prediction from a logistic regression model as: model as:

- A probability of class membership –
- A class assignment by applying threshold to A class assignment, by applying threshold to probability
- threshold re p y resents decision boundar y in feature space

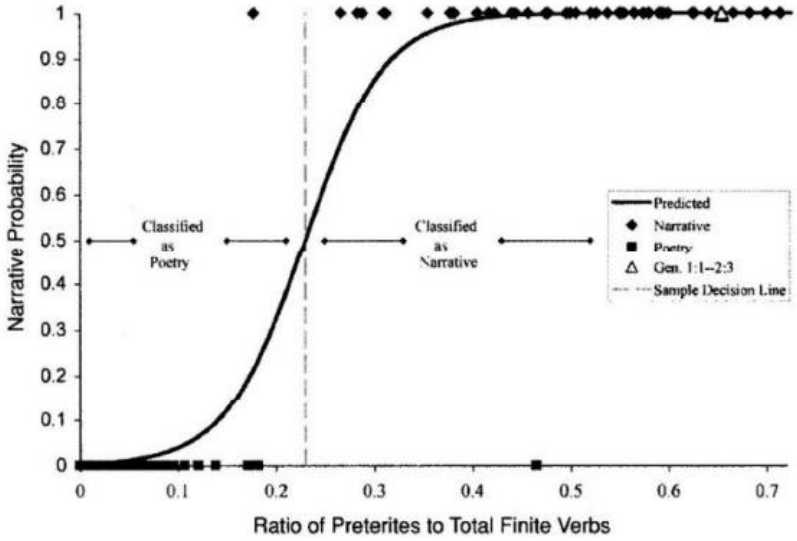
Training a logistic regression model

Need to optimize β so the model gives the best possible reproduction of training set labels possible reproduction of training set labels

– Usually done by numerical approximation of maximum likelihood

– On really large datasets, may use stochastic gradient descent

Logistic regression in one dimension

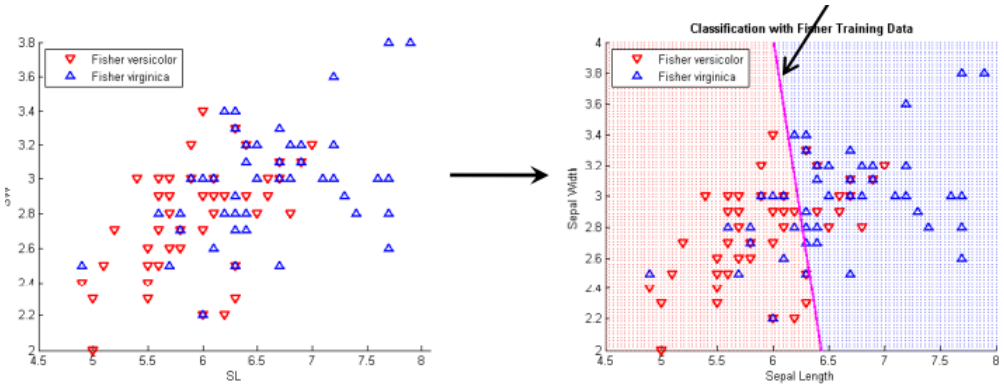


Logistic regression in two dimensions

Subset of Fisher iris dataset

– Two classes –

First two columns (SL, SW)



STEPS:

1) Split the data set into training and test sets.

2) Fit logistic regression model on training set.

3) Prepare Confusion matrix.

4) Compute model accuracy, precision, recall.

Program / Code:-

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as mtp
data_set = pd.read_csv('/content/suv_data.csv')
data_set.head(11)
```

Declaration of Variables:

```
x = data_set.iloc[:,[2,3]].values
y = data_set.iloc[:,4].values
x.shape
y.shape
```

Training Variables:

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size = 0.25, random_state = 0)
y_train.shape
y_test.shape
```

Feature Scaling:

```
from sklearn.preprocessing import StandardScaler
Feature_Scaling = StandardScaler()
Sx_train = Feature_Scaling.fit_transform(x_train)
Sx_test = Feature_Scaling.fit_transform(x_test)
x_test.shape
x_train.shape
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression()
```

Predicting the test set result:

```
y_pred = classifier.predict(x_test)
y_pred
```

Accuracy

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Creation of Confusion Matrix

```
y_pred = classifier.predict(x_test)
y_true = classifier.predict(x_test)
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,y_pred)
cm
```

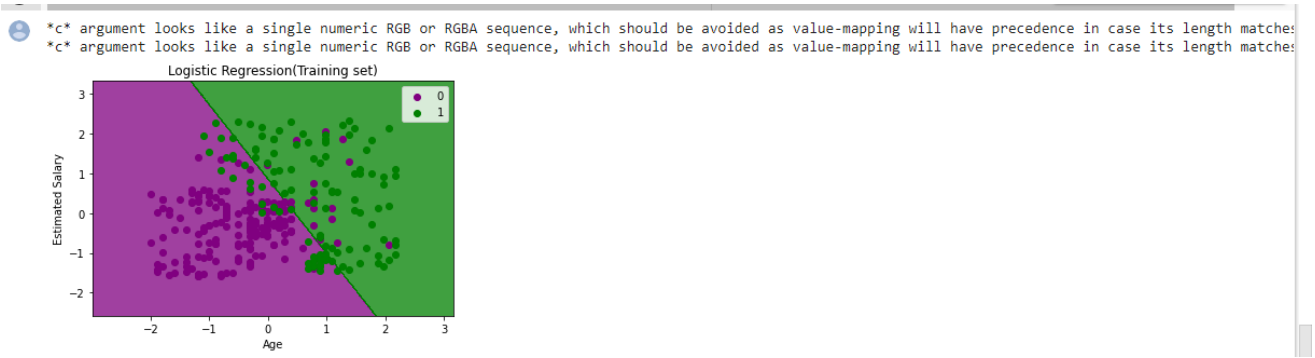
Evaluation of C. M. using Performance Metrics :

```
from sklearn.metrics import classification_report
classification_report(y_test,y_pred)
```

Visualization of Prediction :

```
from matplotlib.colors import ListedColormap
x_set, y_set = x_train, y_train
x1, x2 = np.meshgrid(np.arange(start = x_set[:,0].min() - 1, stop = x_set[:,0].max() +1, step = 0.01),
                     np.arange(start = x_set[:,1].min() - 1,stop=x_set[:,1].max() + 1, step=0.01))
mtp.contourf(x1,x2,classifier.predict(np.array([x1.ravel(),x2.ravel()]).T).reshape(x1.shape),alpha = 0.75, c
map = ListedColormap(('purple','green')))
mtp.xlim(x1.min(),x1.max())
mtp.ylim(x2.min(),x2.max())
for i, j in enumerate(np.unique(y_set)):
  mtp.scatter(x_set[y_set == j, 0],x_set[y_set == j,1],
              c = ListedColormap(('purple','green'))(i), label = j)
mtp.title('Logistic Regression(Training set)')
mtp.xlabel('Age')
mtp.ylabel('Estimated Salary')
mtp.legend()
mtp.show()
```

OUTPUT (SCREENSHOTS)



CONCLUSION: Have understood the concept of Logistic Regression that works on Categorical Data and is used for Classification Purpose. The S curve is generated during the process also known as Sigmoid Function.

DISCUSSION AND VIVA VOCE:

- What is the basic principle of logistic regression?

- What is the practical application of logistic regression?

- What is logistic regression used for in machine learning?

- What are the analytical challenges during model development?

- What are the difference between linear regression and logistic?

REFERENCE:

- https://www.techtarget.com/searchitoperations/definition/virtual-machine-VM
- https://medium.com/analytics-vidhya/20-interview-questions-on-linear-regression-and-logistic-regression-ef4d341d2805
- www.cs.sfu.ca/~han/MachineLearnig.html