

STA 3180 Statistical Modelling: Logistic Regression

Logistic Regression

Definition

Logistic regression is a type of supervised machine learning algorithm used for classification problems. It is a generalized linear model that uses a logistic function to model a binary dependent variable. The goal of logistic regression is to find the best fitting model to describe the relationship between the independent variables and the dependent variable.

Key Concepts

- Logistic regression is used to predict the probability of an event occurring, such as whether a customer will purchase a product or not.
- The logistic function is used to map the input values to output values between 0 and 1.
- Logistic regression can be used for binary classification problems, where the outcome is either a “yes” or “no”.
- Logistic regression can also be used for multi-class classification problems, where the outcome can be one of multiple classes.

Coding Examples

Start of Code

```
# Importing the libraries
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```

```
# Importing the dataset
```

```
dataset = pd.read_csv('Social_Network_Ads.csv')
```

```
X = dataset.iloc[:, [2, 3]].values
```

```
y = dataset.iloc[:, 4].values
```

```
# Splitting the dataset into the Training set and Test set
```

```
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)
```

```
# Feature Scaling
```

```
from sklearn.preprocessing import StandardScaler
```

```
sc = StandardScaler()
```

```
X_train = sc.fit_transform(X_train)
```

```
X_test = sc.transform(X_test)
```

```
# Fitting Logistic Regression to the Training set
```

```
from sklearn.linear_model import LogisticRegression
```

```
classifier = LogisticRegression(random_state = 0)
```

```
classifier.fit(X_train, y_train)
# Predicting the Test set results
y_pred = classifier.predict(X_test)
# Making the Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
### End of Code
```

Practice Multiple Choice Questions

Q1: What is the goal of logistic regression?

- A. To predict the probability of an event occurring
- B. To find the best fitting model to describe the relationship between the independent variables and the dependent variable
- C. To map the input values to output values between 0 and 1
- D. To classify data into different classes

Answer: B. To find the best fitting model to describe the relationship between the independent variables and the dependent variable