

Logistic Regression in R

Spoken Tutorial Project

<https://spoken-tutorial.org>

National Mission on Education through ICT

Yate A Ronald

Debatosh Chakraborty

FOSSEE, IIT Bombay

22 March 2024



Learning Objectives

We will learn about:



Learning Objectives

We will learn about:

- **Logistic Regression**



Learning Objectives

We will learn about:

- **Logistic Regression**
- **Assumptions for Logistic Regression**



Learning Objectives

We will learn about:

- **Logistic Regression**
- **Assumptions for Logistic Regression**
- **Advantages of Logistic Regression**



Learning Objectives

We will learn about:

- **Logistic Regression**
- **Assumptions for Logistic Regression**
- **Advantages of Logistic Regression**
- **Implementation of Logistic Regression in R using Raisin Dataset**



Learning Objectives

- **Model Evaluation**



Learning Objectives

- **Model Evaluation**
- **Visualization of the model Decision Boundary**



Learning Objectives

- **Model Evaluation**
- **Visualization of the model Decision Boundary**
- **Limitations of Logistic Regression**



System Specifications



System Specifications

- Windows 11



System Specifications

- Windows 11
- R v 4.3.0



System Specifications

- Windows 11
- R v 4.3.0
- RStudio v 2023.06.1



System Specifications

- Windows 11
- R v 4.3.0
- RStudio v 2023.06.1



System Specifications

- Windows 11
- R v 4.3.0
- RStudio v 2023.06.1

It is recommended to install R version 4.2.0 or higher



Pre-requisites



Pre-requisites

To follow this tutorial, the learner should know:



Pre-requisites

To follow this tutorial, the learner should know:

- **Basic Programming in R**



Pre-requisites

To follow this tutorial, the learner should know:

- **Basic Programming in R**
- **Basics of Machine Learning**



Pre-requisites

To follow this tutorial, the learner should know:

- Basic Programming in R
- Basics of Machine Learning
- If not, please access the relevant tutorials on R on this website
<https://spoken-tutorial.org>



Logistic Regression



Logistic Regression

- **Logistic Regression is a statistical method used for classification**



Logistic Regression

- **Logistic Regression is a statistical method used for classification**
- **It models the probability of success for the explanatory variable**



Logistic Regression

- It predicts the probability, unlike the response in linear regression



Logistic Regression

- It predicts the probability, unlike the response in linear regression
- The predicted probability is used as a classifier



Logistic Regression

- It predicts the probability, unlike the response in linear regression
- The predicted probability is used as a classifier
- The probability of success is modeled using the logit (log odds) function



About Logistic Regression classifier

- It is a linear classifier, as the logistic regression model has a linear logit



About Logistic Regression classifier

- It is a linear classifier, as the logistic regression model has a linear logit
- It is often used when the response variable is categorical



Assumptions for Logistic Regression

- The distribution of the dependent variable is Bernoulli
- The data records are independent



Advantages of Logistic Regression

- It provides estimates of regression coefficients along with their standard errors
- It doesn't need explanatory variables to be necessarily continuous
- In this sense, it is a more general classifier than LDA and QDA



Implementation of Logistic Regression



Implementation of Logistic Regression

- Now let us implement Logistic Regression on the raisin dataset with two chosen variables



Implementation of Logistic Regression

- Now let us implement Logistic Regression on the raisin dataset with two chosen variables
- For more information on Raisin data please see the Additional Reading material on this tutorial page



Download Files

We will use:



Download Files

We will use:

- A script file **LogisticRegression.R**



Download Files

We will use:

- A script file **LogisticRegression.R**
- Raisin Dataset 'raisin.xlsx'



Download Files

We will use:

- A script file **LogisticRegression.R**
- Raisin Dataset 'raisin.xlsx'



Download Files

We will use:

- A script file **LogisticRegression.R**
- Raisin Dataset 'raisin.xlsx'

Please download these files from the **Code files** link of this tutorial

Make a copy and then use them while practising



Limitations of Logistic Regression

- It is sensitive to outliers
- It may underperform when explanatory variables are multicollinear



Summary

In this tutorial we have learnt about:

- **Logistic Regression**
- **Assumptions for Logistic Regression**
- **Advantages of Logistic Regression**
- **Implementation of Logistic Regression using Raisin Dataset**



Summary

- **Model Evaluation**
- **Visualization of the model Decision Boundary**
- **Limitations of Logistic Regression model**



Assignment

- Apply Logistic Regression on the wine dataset
- This dataset can be found in the **HDclassif** package
- Install the package and import the dataset using the **data()** command
- Measure the accuracy of the model



About the Spoken Tutorial Project

- Watch the video available at https://spoken-tutorial.org/What_is_a_Spoken_Tutorial
- It summarises the Spoken Tutorial project



About the Spoken Tutorial Project

- Watch the video available at https://spoken-tutorial.org/What_is_a_Spoken_Tutorial
- It summarises the Spoken Tutorial project
- If you do not have good bandwidth, you can download and watch it



Spoken Tutorial Workshops

The Spoken Tutorial Project Team

- Conducts workshops using spoken tutorials
- Gives certificates to those who pass an online test
- For more details, please write to contact@spoken-tutorial.org



Answers for THIS Spoken Tutorial

- Questions in THIS Spoken Tutorial?
- Visit <https://forums.spoken-tutorial.org>
- Choose the minute and second where you have the question
- Explain your question briefly
- The Spoken Tutorial project will ensure an answer



Forum to answer questions

- Do you have any general/technical questions?
- Please visit the forum given in the link
<https://forums.fossee.in/>



Textbook Companion Project

- The FOSSEE team coordinates the coding of solved examples of popular books and case study projects
- We give certificates to those who do this
- For more details, please visit these sites

<https://r.fossee.in/>

<https://fossee.in/>



Acknowledgements

- **Spoken Tutorial project was established by the Ministry of Education(MoE), Govt of India**



Thank You

- This tutorial is contributed by Yate A Ronald O and Debatosh Chakraborty from IIT Bombay
- Thank you for joining

