

Linear Discriminant Analysis in R

Spoken Tutorial Project

<https://spoken-tutorial.org>

National Mission on Education through ICT

<https://sakshat.ac.in>

Tanmay Srinath

Madhuri Ganapathi

IIT Bombay

24 January 2022



Learning Objectives

We will learn about:



Learning Objectives

We will learn about:

- ▶ **Linear discriminant analysis or LDA**



Learning Objectives

We will learn about:

- ▶ **Linear discriminant analysis or LDA**
- ▶ **Applications of LDA**



Learning Objectives

We will learn about:

- ▶ **Linear discriminant analysis or LDA**
- ▶ **Applications of LDA**
- ▶ **Assumptions of LDA**



Learning Objectives

We will learn about:

- ▶ **Linear discriminant analysis or LDA**
- ▶ **Applications of LDA**
- ▶ **Assumptions of LDA**
- ▶ **Robustness of LDA**



Learning Objectives

We will learn about:

- ▶ **Linear discriminant analysis or LDA**
- ▶ **Applications of LDA**
- ▶ **Assumptions of LDA**
- ▶ **Robustness of LDA**
- ▶ **LDA on iris dataset**



System Specifications



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**
- ▶ **R version 4.1.2**



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**
- ▶ **R version 4.1.2**
- ▶ **RStudio version 1.4.1717**



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**
- ▶ **R version 4.1.2**
- ▶ **RStudio version 1.4.1717**



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**
- ▶ **R version 4.1.2**
- ▶ **RStudio version 1.4.1717**

R version 4.1.0 or higher



Pre-requisites



Pre-requisites

► Basics of R Programming



Pre-requisites

- ▶ **Basics of R Programming**
- ▶ **Basics of Machine Learning using R**



Pre-requisites

- ▶ **Basics of R Programming**
- ▶ **Basics of Machine Learning using R**



Pre-requisites

- ▶ Basics of R Programming
- ▶ Basics of Machine Learning using R

If not, please access the relevant tutorials on R on

<https://spoken-tutorial.org/>



Linear Discriminant Analysis



Linear Discriminant Analysis

- ▶ It is a linear combination of features that separates two or more classes



Linear Discriminant Analysis

- ▶ It is a linear combination of features that separates two or more classes
- ▶ It is the best classifier when data is gaussian



Linear Discriminant Analysis

- ▶ It has the smallest misclassification error



Linear Discriminant Analysis

- ▶ It has the smallest misclassification error
- ▶ It relies on the Bayes Classifier



Applications of LDA



Applications of LDA

- ▶ LDA is primarily a multi-class classifier



Assumptions of LDA



Assumptions of LDA

- ▶ **Multivariate normality:** The records are gaussian



Assumptions of LDA

- ▶ **Multivariate normality:** The records are gaussian
- ▶ **Homoscedasticity:** Variance and covariance structure among classes are the same



Assumptions of LDA

- ▶ **Multivariate normality:** The records are gaussian
- ▶ **Homoscedasticity:** Variance and covariance structure among classes are the same
- ▶ **Independence:** The records are independent of each other



Robustness of LDA



Robustness of LDA

- ▶ LDA is robust to slight violations of these assumptions



Robustness of LDA

- ▶ LDA is robust to slight violations of these assumptions
- ▶ If data is heteroscedastic, LDA becomes QDA



Robustness of LDA

- ▶ LDA is robust to slight violations of these assumptions
- ▶ If data is heteroscedastic, LDA becomes QDA
- ▶ QDA can be performed when the covariance structure of the classes is different



Implement **LDA** on the iris dataset



Download Files

We will use:



Download Files

We will use:

- ▶ A script file LDA.R



Download Files

We will use:

- ▶ A script file **LDA.R**



Download Files

We will use:

- ▶ A script file **LDA.R**

Download this file from the **Code files** link of this tutorial

Make a copy and then use it for practising



LDA Classifier Model

We will create a **LDA classifier model** on the iris dataset



Summary

In this tutorial we have learnt about:

- ▶ **Linear discriminant analysis or LDA**
- ▶ **Applications of LDA**
- ▶ **Assumptions of LDA**
- ▶ **Robustness of LDA**
- ▶ **LDA on `iris` dataset**



Assignment



Assignment

- ▶ **Perform LDA on inbuilt PlantGrowth dataset**



Assignment

- ▶ **Perform LDA on inbuilt `PlantGrowth` dataset**
- ▶ **Evaluate the model using a confusion matrix and visualise the results**



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 FOSSEE project will ensure an answer**

You will have to register to ask questions



Forum to answer questions

- ▶ Questions not related to the Spoken Tutorial?
- ▶ Do you have general / technical questions on the Software?
- ▶ Please visit the FOSSEE Forum
<https://forums.fossee.in/>
- ▶ Choose the Software and post your question



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

- ▶ **The Spoken Tutorial and FOSSEE projects are funded by the Ministry of Education, Govt. of India**



About the Contributors

- ▶ **This tutorial is contributed by Tanmay Srinath and Madhuri Ganapathi, IIT Bombay**

