

# Introduction to Machine Learning in R

Spoken Tutorial Project

<https://spoken-tutorial.org>

National Mission on Education through ICT

Debatosh Chakraborty

FOSSEE, IIT Bombay

24 January 2024



# Learning Objectives

**We will learn about:**



# Learning Objectives

**We will learn about:**

- **Machine Learning**



# Learning Objectives

**We will learn about:**

- **Machine Learning**
- **Supervised and Unsupervised Learning**



# Learning Objectives

**We will learn about:**

- **Machine Learning**
- **Supervised and Unsupervised Learning**
- **Workflow of ML Classifier Algorithm**



# Learning Objectives

- Visualizing Feature Space



# Learning Objectives

- Visualizing Feature Space
- Constructing a dummy classifier



# Learning Objectives

- Visualizing Feature Space
- Constructing a dummy classifier
- Evaluation of the chosen dummy classifier





# 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
- To use **GGPlot2** and **dplyr** package



# Pre-requisites

To follow this tutorial, the learner should know:

- Basic Programming in R
- To use GGPlot2 and dplyr package
- If not, please access the relevant tutorials on R on this website

<https://spoken-tutorial.org>



# Machine Learning



# Machine Learning

- ML enables computers to learn from data



# Machine Learning

- ML enables computers to learn from data
- ML algorithms automate the learning process from data through patterns



# Machine Learning

- ML enables computers to learn from data
- ML algorithms automate the learning process from data through patterns
- Their primary role is prediction, classification or clustering of data



# Machine Learning

- ML algorithms are applied in several applications





# Machine Learning

- ML algorithms are applied in several applications
- For example Natural Language Processing, Image and speech recognition, etc



# Types of Machine Learning

**ML algorithms include the following types and tasks:**



# Types of Machine Learning

ML algorithms include the following types and tasks:

- **Supervised learning: Prediction and Classification**



# Types of Machine Learning

ML algorithms include the following types and tasks:

- Supervised learning: Prediction and Classification
- Unsupervised learning: Clustering



# Types of Machine Learning

- **Semi-supervised learning**



# Types of Machine Learning

- Semi-supervised learning
- Reinforcement learning



# Types of Machine Learning

- **Semi-supervised learning**
- **Reinforcement learning**



# Types of Machine Learning

- Semi-supervised learning
- Reinforcement learning

**In this series, we will focus on Supervised and Unsupervised learning algorithms**





# Supervised and Unsupervised Learning

## Supervised learning: Labeled data

- ML algorithms predict labels for unseen features



# Supervised and Unsupervised Learning

## Supervised learning: Labeled data

- ML algorithms predict labels for unseen features
- They predict based on given features and labels of data



# Supervised and Unsupervised Learning

## Unsupervised learning: Unlabeled data

- ML algorithms develop a mechanism to group similar features into clusters



# Supervised and Unsupervised Learning

## Unsupervised learning: Unlabeled data

- ML algorithms develop a mechanism to group similar features into clusters
- And label them for future analysis



# Classification and Regression

- Supervised learning consists of Regression and Classification



# Classification and Regression

- Supervised learning consists of Regression and Classification
- Regression is applied to predict and learn continuous-valued responses from features



# Classification and Regression

- Supervised learning consists of Regression and Classification
- Regression is applied to predict and learn continuous-valued responses from features
- Regression techniques include Linear, Spline, Ridge, Lasso, and others



# Classification and Regression

- **Classification is applied to predict the class of a discrete (labeled) response from features**





# Classification and Regression

- **Classification is applied to predict the class of a discrete (labeled) response from features**
- **Classification techniques include Logistic Regression, Decision Tree, SVM, and others**



# Workflow of an ML Classifier algorithm

- **Feature Space:** Collection of all possible values of the features



# Workflow of an ML Classifier algorithm

- **Feature Space: Collection of all possible values of the features**
- **A classification algorithm partitions the feature space into multiple classes**



# Workflow of an ML Classifier algorithm

- **Feature Space:** Collection of all possible values of the features
- A classification algorithm partitions the feature space into multiple classes
- Data is split into training and testing sets to learn and evaluate the algorithm



# Workflow of an ML Classifier algorithm

- The model learns from the training data to create partitions of feature space



# Workflow of an ML Classifier algorithm

- The model learns from the training data to create partitions of feature space
- The model is evaluated on the test dataset through performance metrics



# Dataset



# Dataset

- Let's use Raisin dataset with two chosen variables/features to understand a classification problem





# Dataset

- Let's use Raisin dataset with two chosen variables/features to understand a classification problem
- For more information on Raisin data please refer to [Additional Reading Material](#) on this tutorial page



# Download Files

**We will use:**



# Download Files

**We will use:**

- **A script file Intro.R**



# Download Files

**We will use:**

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



# Download Files

We will use:

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



# Download Files

We will use:

- A script file **Intro.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 practicing



# Summary

**In this tutorial we have learnt about:**

- **Machine Learning**
- **Supervised and Unsupervised Learning**
- **Workflow of ML Classifier Algorithm**



# Summary

- **Visualizing Feature Space**
- **Constructing a dummy classifier**
- **Evaluation of the chosen dummy classifier**





# Assignment

- Use a vertical line as a classifier to partition the feature space
- Plot the decision boundary for the same
- Evaluate the classifier on the test dataset



# About the Spoken Tutorial Project

- Watch the video available at [https://spoken-tutorial.org/What\\_is\\_a\\_Spoken\\_Tutorial](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](mailto: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/>



# R activities

- The FOSSEE team coordinates the Textbook Companion, Lab Migration and the 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 **Debatosh Chakraborty** from **IIT Bombay**
- Thank you for joining

