Supervised Learning

Spoken Tutorial Project
https://spoken-tutorial.org
National Mission on Education through ICT
https://sakshat.ac.in

Sudhakar Kumar
Madhuri Ganapathi
IIT Bombay

28 December 2021

Learning Objectives

We will learn about:

Learning Objectives

We will learn about:

- `Machine Learning` **and its types**

Learning Objectives

We will learn about:

- `Machine Learning` **and its types**
- `Supervised` **learning**

Learning Objectives

We will learn about:

- `Machine Learning` **and its types**
- `Supervised` **learning**
- **Classification model on** `iris` **data**

Learning Objectives

We will learn about:

- `Machine Learning` **and its types**
- `Supervised` **learning**
- **Classification model on** `iris` **data**
- **Confusion matrix**

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**

Install R version 4.1.0 or higher

Prerequisites

Prerequisites

- **Basics of R programming**

Prerequisites

- **Basics of R programming**
- **Basics of Statistics**

Prerequisites

- **Basics of R programming**
- **Basics of Statistics**

If not, please access the relevant tutorials on `R` on

https://spoken-tutorial.org/

What is Machine Learning?

What is Machine Learning?

- **ML is a science that enables computers to learn without being explicitly programmed**

What is Machine Learning?

- ML is a science that enables computers to learn without being explicitly programmed
- Its applications include self-driven cars, speech recognition, etc

What is Machine Learning?

- **ML is a science that enables computers to learn without being explicitly programmed**
- **Its applications include self-driven cars, speech recognition, etc**
- **It is seen as a subset of** `Artificial Intelligence`

Classification of Machine Learning

Classification of Machine Learning

ML is broadly classified into the following types:

Classification of Machine Learning

ML is broadly classified into the following types:

- `Supervised` **learning**

Classification of Machine Learning

ML is broadly classified into the following types:

- `Supervised` **learning**
- `Unsupervised` **learning**

Classification of Machine Learning

ML is broadly classified into the following types:

- `Supervised` **learning**
- `Unsupervised` **learning**
- `Semi-supervised` **learning**

Classification of Machine Learning

ML is broadly classified into the following types:

- `Supervised` **learning**
- `Unsupervised` **learning**
- `Semi-supervised` **learning**
- `Reinforcement` **learning**

Iris Flower

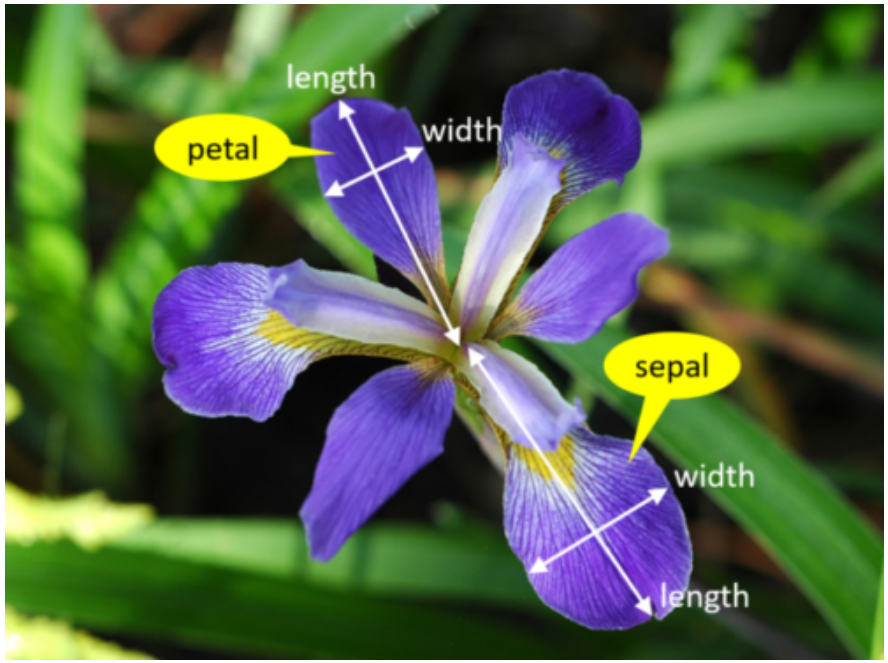

Figure: **An iris flower with its parameters**
`https://www.integratedots.com/`

Iris Flower

There are two critical parameters of an `iris` flower:

Iris Flower

There are two critical parameters of an `iris` flower:

- **Sepal**

Iris Flower

There are two critical parameters of an `iris` flower:

- **Sepal**
- **Petal**

Iris Flower

There are two critical parameters of an `iris` flower:

- **Sepal**
- **Petal**

Iris Flower

There are two critical parameters of an `iris` flower:

- **Sepal**
- **Petal**

One can measure the length and width of these two parameters

Species of an Iris Flower

Three different species of an `iris` flower

Iris setosa | *Iris versicolor* | *Iris virginica*

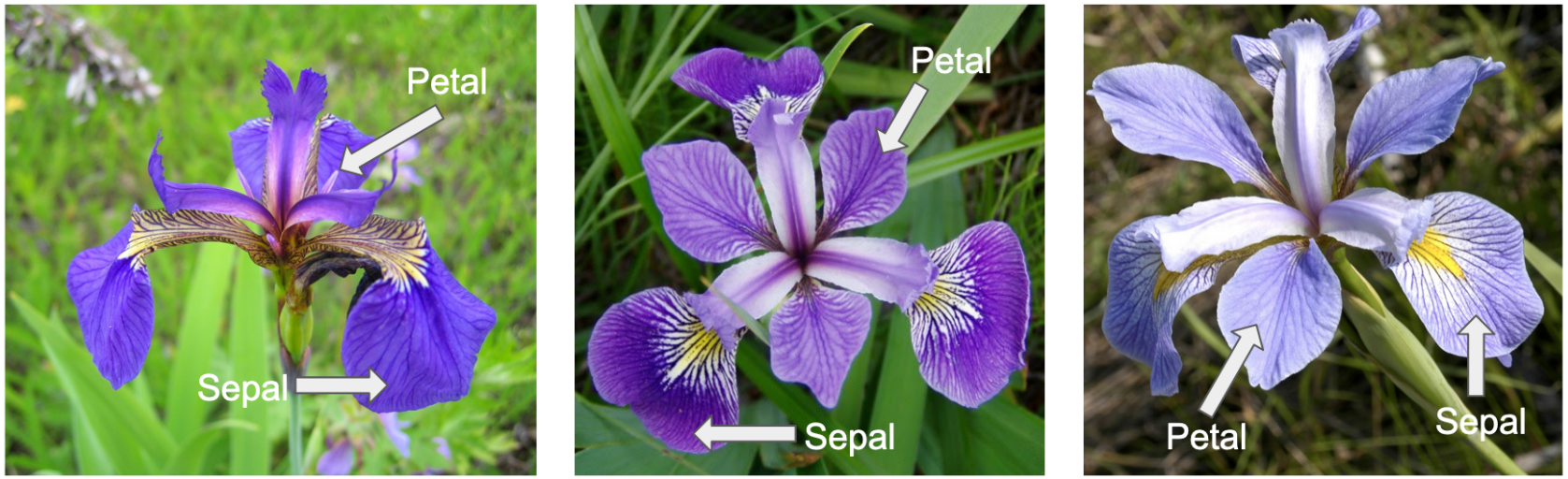

Figure: **Species of an iris flower**

`https://towardsdatascience.com/`

Tabulating the Data

Tabulating the Data

- **A botanist wants to distinguish the species of `iris` flowers**

Tabulating the Data

- **A botanist wants to distinguish the species of `iris` flowers**
- **She collects the four features of some `iris` flowers**

Tabulating the Data

- **A botanist wants to distinguish the species of `iris` flowers**
- **She collects the four features of some `iris` flowers**
  - `Sepal` **length (S_l),** `Sepal` **width (S_w)**

Tabulating the Data

- **A botanist wants to distinguish the species of `iris` flowers**
- **She collects the four features of some `iris` flowers**
  - `Sepal` **length** (S_l), `Sepal` **width** (S_w)
  - `Petal` **length** (P_l), `Petal` **width** (P_w)

Tabulating the Data

- **She gets these flowers labeled as one of the three species by an expert**

Download Files

Download Files

For this tutorial, we will use,

Download Files

For this tutorial, we will use,

- **A data set** `iris.csv`

Download Files

For this tutorial, we will use,

- **A data set** `iris.csv`
- **A script file** `irisModel.R`

Download Files

For this tutorial, we will use,

- **A data set** `iris.csv`
- **A script file** `irisModel.R`

Download Files

For this tutorial, we will use,

- **A data set** `iris.csv`
- **A script file** `irisModel.R`

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

Make a copy and then use them for practising

Posing the Problem

Posing the Problem

Suppose that the botanist considers the following:

Posing the Problem

Suppose that the botanist considers the following:

- **Can I build a model that learns from labels of known species?**

Posing the Problem

Suppose that the botanist considers the following:

- **Can I build a model that learns from labels of known species?**
- **Can this model accurately predict the species from its measurements?**

Mapping of Features and Label

Mapping of Features and Label

- **We will map the dimensions of** `sepal` **and** `petal` **to** `iris` **species**

Mapping of Features and Label

- **We will map the dimensions of `sepal` and `petal` to `iris` species**
- **The classification model would work as a function as given below:**
  $f(S_l, S_w, P_l, P_w) \rightarrow Species$

Mapping of Features and Label

- **We will map the dimensions of `sepal` and `petal` to `iris` species**
- **The classification model would work as a function as given below:**
  $f(S_l, S_w, P_l, P_w) \rightarrow Species$
- **This mechanism is `supervised` learning**

Supervised Learning

In supervised learning,

Supervised Learning

In supervised learning,

- **The desired output labels are available for training datasets**

Supervised Learning

In supervised learning,

- **The desired output labels are available for training datasets**
- **These labels can be called** `supervisors`

Supervised Learning

- **While learning, the model makes predictions using the given training dataset**

Supervised Learning

- **While learning, the model makes predictions using the given training dataset**
- **The model iteratively makes predictions on the training dataset**

Supervised Learning

- **While learning, the model makes predictions using the given training dataset**
- **The model iteratively makes predictions on the training dataset**
- **The supervisor corrects the model**

Types of Supervised Learning

There are two types of supervised learning:

Types of Supervised Learning

There are two types of supervised learning:

- **Regression and Classification**

Types of Supervised Learning

- `Regression`: **Applied to predict a continuous-valued output**

Types of Supervised Learning

- `Regression`: **Applied to predict a continuous-valued output**
- **For example, predicting prices for the real estate sector**

Types of Supervised Learning

- `Classification`: **Applied to predict a discrete-valued output**

Types of Supervised Learning

- `Classification`: **Applied to predict a discrete-valued output**
- **For example, predicting the species of an iris flower**

Confusion Matrix

Confusion Matrix

- **It is a performance measurement for `ML` classification problems**

Confusion Matrix

- **It is a performance measurement for ML classification problems**
- **In these classification problems, the output can be two or more classes**

Summary

In this tutorial, we have learnt about,

- `Machine Learning` **and its types**
- `Supervised` **learning**
- **Classification model on** `iris` **data**
- **Confusion matrix**

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 for specific questions

- **The Spoken Tutorial forum is for specific questions on this tutorial**
- **Please do not post unrelated and general questions on them**
- **This will help reduce the clutter**
- **With less clutter, we can use these discussions as instructional material**

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 Sudhakar Kumar and Madhuri Ganapathi, IIT Bombay**