



# Piscine Pro AI / Machine Learning

## Multinomial Logistic Regression

*Summary: In this Module, you will learn about Multinomial Logistic Regression.*

*Version: 1.00*

# Contents

<b>I</b>	<b>Introduction</b>	<b>2</b>
<b>II</b>	<b>General instructions</b>	<b>3</b>
<b>III</b>	<b>Exercise 00</b>	<b>4</b>
<b>IV</b>	<b>Exercise 01</b>	<b>5</b>
<b>V</b>	<b>Exercise 02</b>	<b>6</b>
<b>VI</b>	<b>Submission and peer-evaluation</b>	<b>7</b>

# Chapter I

## Introduction

Greetings!



If you haven't already done so, read `en.toolkit.pdf`.

What this Module will cover:

Welcome to this module, where we will dive into the realm of multinomial logistic regression for tackling multiclass classification tasks. Building on your foundational understanding of logistic regression, we will delve into the intricacies of predicting outcomes across multiple classes. This versatile technique extends logistic regression to scenarios where there are more than two classes to predict. Through hands-on exercises, you will uncover the concepts and methodologies underlying multinomial logistic regression.

In this module, we will be using anonymized data, and you will need to understand it.

Wishing you success in your learning journey.

# Chapter II


## General instructions

Unless explicitly specified, the following rules will apply every day of this Piscine Pro.

- This subject is the one and only trustable source. Don't trust any rumor.
- This subject can be updated up to one hour before the turn-in deadline.
- The assignments in a subject must be done in the given order. Later assignments won't be rated unless all the previous ones are perfectly executed.
- Be careful about the access rights of your files and folders.
- Your assignments will be evaluated by your peers.
- You must not leave in your turn-in your workspace any file other than the ones explicitly requested By the assignments.
- You have a question? Ask your left neighbor. Otherwise, try your luck with your right neighbor.
- Every technical answer you might need is available in the **man** or on the Internet.
- By Thor, by Odin! Use your brain!!!

# Chapter III


## Exercise 00

	Exercise 00
Data exploration	
Turn-in directory : <i>ex00/</i>	
Files to turn in : <b>Medium01.ipynb</b>	
Allowed functions : <b>A11</b>	

For this first exercise, you'll need to load the data into colab and do some data exploration, i.e. you'll need to understand your data.

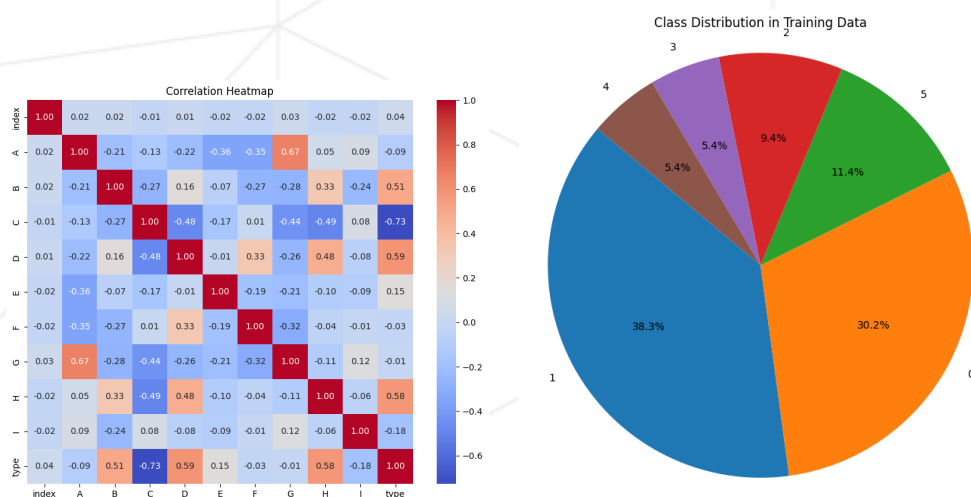
# Chapter IV

## Exercise 01

	Exercise 01
Data Visualisation	
Turn-in directory : <i>ex01/</i>	
Files to turn in : <b>Medium01.ipynb</b>	
Allowed functions : <b>All</b>	


In this exercise, you'll be working with anonymized data. The objective is to create visual representations that depict the relationships between different numerical attributes and certain outcomes. By generating box plots for each numerical characteristic in relation to these outcomes and creating a heat map to visualize the correlations between these characteristics, you can derive an understanding of how these anonymized numerical attributes might influence certain results.

You should have something like this:



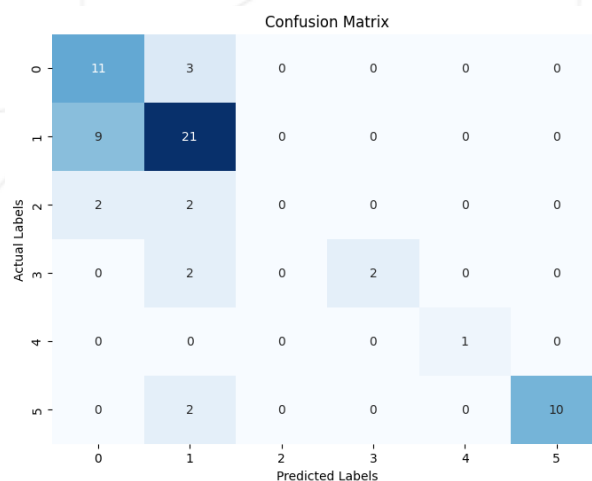
# Chapter V

## Exercise 02

	Exercise 02
Model	
Turn-in directory : <i>ex02/</i>	
Files to turn in : <b>Medium01.ipynb</b>	
Allowed functions : <b>A11</b>	

In this exercise, you'll delve into the field of polytomous logistic regression, which is used for classifying data into multiple categories. Make predictions, assess its performance using techniques like the confusion matrix, and present the results visually using a heatmap.

You should have something like this:



You must have an accuracy greater than 0.6 with your test set.

# Chapter VI

## Submission and peer-evaluation

- Create a `professional_training_medium` folder at the root of your home, and move around in it.
- Create a new `module01` folder and navigate to it.



Please note, during your defense anything that is not present in the folder for the day will not be checked.