



# Piscine Pro AI / Machine Learning

## Binary Classification with Logistic Regression

*Summary: In this Module, you will learn about 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>Bonus part</b>	<b>7</b>
<b>VII</b>	<b>Submission and peer-evaluation</b>	<b>8</b>

# Chapter I

## Introduction

Greetings!



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

What this Module will cover:

In this module, we will venture into the realm of binary classification using logistic regression. Building upon the foundational understanding of linear regression, we will explore how logistic regression is a powerful tool for predicting binary outcomes. We will unravel the concepts and methodologies behind logistic regression. By employing key libraries such as scikit-learn for model implementation, pandas for data manipulation, and data visualization tools, we will unlock the ability to effectively model and predict binary outcomes. Logistic regression enables us to delve into the intricacies of classification, bridging the gap between linear regression and classification tasks. By the end of this module, you will possess the skills to confidently navigate binary classification problems, and you'll be well-versed in leveraging logistic regression as a valuable tool in your data analysis toolkit.

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>Medium00.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.

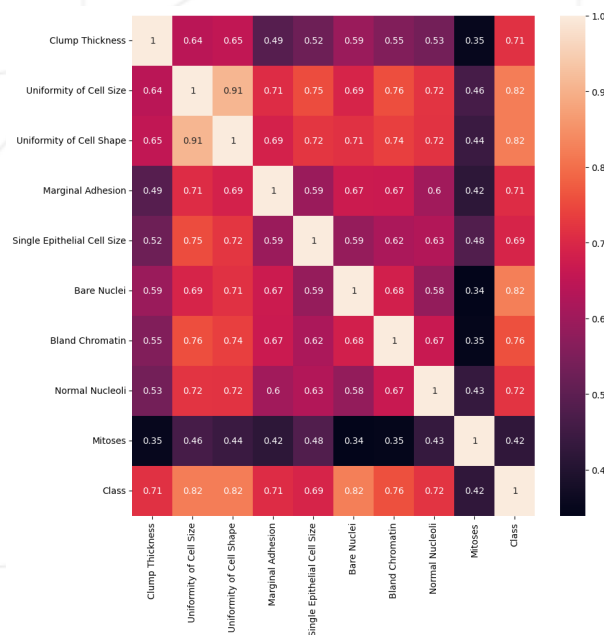
# Chapter IV

## Exercise 01

	Exercise 01
Data Visualisation	
Turn-in directory : <i>ex01/</i>	
Files to turn in : <b>Medium00.ipynb</b>	
Allowed functions : <b>All</b>	


In this exercise, you'll visualize the relationship between the numerical attributes of cancer data and their potential impact on cancer outcomes. By creating box plots for each numerical characteristic in relation to cancer outcomes, and visualizing the correlation between these characteristics using a heat map, you can understand how different numerical attributes can influence cancer prognosis.

You should have something like this:



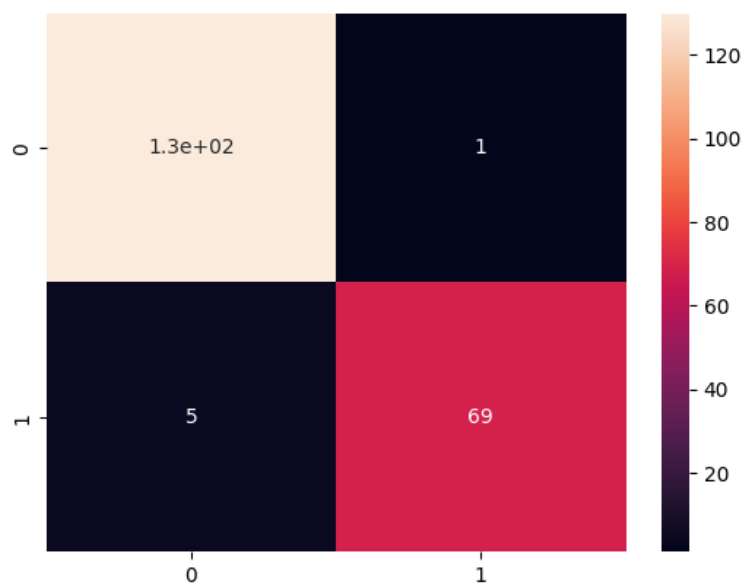
# Chapter V

## Exercise 02

	Exercise 02
Model	
Turn-in directory : <i>ex02/</i>	
Files to turn in : <b>Medium00.ipynb</b>	
Allowed functions : <b>All</b>	

In this exercise, you'll dive into the realm of logistic regression for classification tasks. You'll learn to build a logistic regression model, make predictions, evaluate its performance using the confusion matrix, and visualize the results using a heatmap. The dataset at hand involves attributes and outcomes related to a classification problem.

You should have something like this:



You must have an accuracy greater than 0.9 with your test set.

# Chapter VI

## Bonus part

If you have some time, you can try to validate this little bonus. It is not mandatory, so don't waste too much time doing it.

You'll have to redo this whole module without the libraries, you can start with the simple parts like the confusion matrix and classification report, look for the mathematical formulas on the internet.



The bonus part will only be assessed if the mandatory part is PERFECT. Perfect means the mandatory part has been integrally done and works without malfunctioning. If you have not passed ALL the mandatory requirements, your bonus part will not be evaluated at all.



# Chapter VII

## Submission and peer-evaluation

- Create a `professional_training_medium` folder at the root of your home, and move around in it.
- Create a new `module00` 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.