

**COMPUTING SUBJECT:** Machine Learning

**TYPE:** WORK ASSIGNMENT

**IDENTIFICATION:** Logistic Regression Iris

**COPYRIGHT:** *Michael Claudius*

**DEGREE OF DIFFICULTY:** Medium

**TIME CONSUMPTION:** 1 hours

**EXTENT:** < 150 lines

**OBJECTIVE:** Basic understanding of binary classification.  
MNIST data set

**COMMANDS:**

## **IDENTIFICATION:** LogisticRegIris/MICL

### The Mission

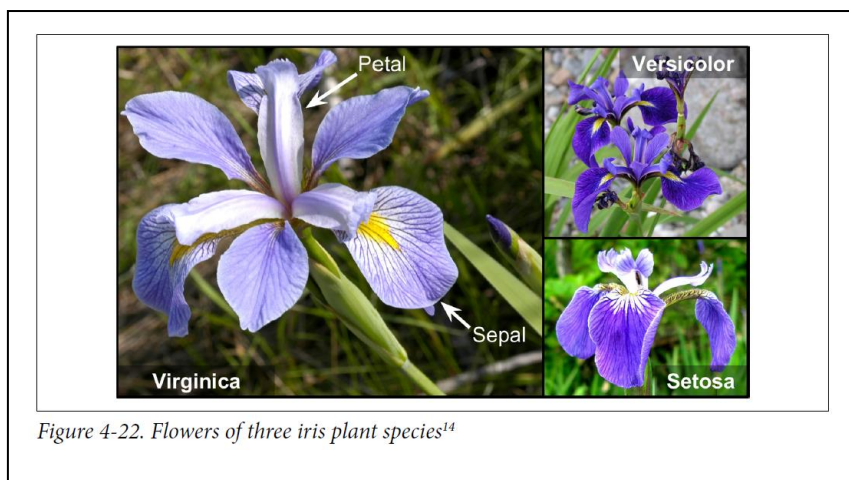
To understand the idea behind logistic regression and decision boundaries.

### Precondition

You must have read the first part of Training Models in chapter 4.

### The problem

Given a data set of 150 Iris flowers (X) and the label, the correct value (Y), you are to train a logistic regression classification both binary and multinomial with 3 classes. You are to use the Iris data set built in Sklearn.



### Useful links

<https://matplotlib.org/3.1.0/tutorials/introductory/pyplot.html>

### Assignment 1: Download data set and program

As the data set is built in we don't need to download it.

For the program it is **easiest** to download a downsized version with only the necessary code for logistic regression from your teacher's home page. The file is "*LogisticRegressionCh4.ipynb*" encapsulated in a .zip file.

Alternatively You could choose to from the Github to copy the Chapter 4 Jupyter program, "*04-training-linear-models.ipynb*", into the folder for your solutions (Machine Learning/Solutions) folder. Rename it "*MiniLogistic.ipynb*", before you cut away a lot of not-necessary code.

### Assignment 2: Application program, adjusting the program

Start Jupyter and open the file. You will now delete and out-comment some lines/blocks.

First, scroll down to the heading "**Exercise solutions**" app. Cell [17].

**Exercise solutions**

---

**1. to 11.**

See appendix A.

**Extra material**

**Dummy (ie. random) classifier**

Delete all cells from this cell and cell [17] and down to the end. They are superfluous for this assignment.

*Now we can start to execute the cells.*

### Assignment 3: Logistic regression classifier

Run the cells one by one and on the way discuss the topics and write down the answers to the following questions:

- a. What is a sigmoid function good for ?
- b. Discuss the DESCR attributes.
- c. Print out the full data set or a part of it. (Cell [6]). What do you see ?
- d. Print out the full target set. What do you see ?
- e. Try to change some of the input parameters for predict- function.  
Look at the different probability values for the classes.  
Do it in the following way:  
Add more cells and  
    Use the petal data: [5, 2] as input parameters for predict- function  
    Look at the different probability values for the classes  
    Use the petal data: [5, 1.5] as input parameters for predict- function  
    Look at the different probability values for the classes.  
Write down the data you will use it later in Assignment 5!!
- f. How to declare binary LogisticRegression ? (Show the code)
- g. How to declare Softmax? (Show the code)  
And discuss the parameters “solver” and “C”  
Tip: [https://en.wikipedia.org/wiki/Limited-memory\\_BFGS](https://en.wikipedia.org/wiki/Limited-memory_BFGS)
- h. Go to the homepage for ScikitLearn.org and see the definition of LogisticRegression constructor. There are many parameters.  
Compare the parameters with the one used in the program code and explain “C” and 2 others of them.  
Tip:  
[https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

### Assignment 4: Using sepal data

Lets see how it looks using only setal data.

At the bottom of the program try to add cells using only Softmax and change it to use setal-data.

So you copy and paste the cells and then change the data to be used and also the values of data to be predicted.

Tip: `X = iris["data"][:, (0, 1)]`

Tip: Use [5,2] and [7,2]

Remember to write down the probabilities.

### Assignment 5: Using petal data and septal data

You are now to investigate how much better will the logistic regression perform IF it use 4 features:

-petal length, petal width, septal length, petal width

Do you still want to plot decision boundaries ?

- Yes: Why?
- No: Why not?

Now use Softmax on full data features and tune *predict* and *predict\_proba* functions on combinations of the data set already investigated. For example:

[5, 2, 5, 2]

[5, 1.5, 7, 2]

***Congratulation: You are now an expert-level 1 in logistic regression 😊***