Project 2

Kishan Sarpangala

Question:

The Iris is a four dimension and three class data.

**(1)** Follow the code in hands-on Chapter 4 to perform logistic regression (**select one of the three class as positive**)

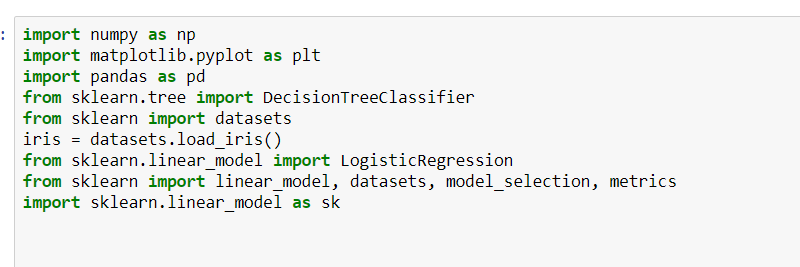
**(2)** Softmax regression for all three classes but display a decision boundary diagram using the two other dimensions not shown in Figure 4-25.

Answer:

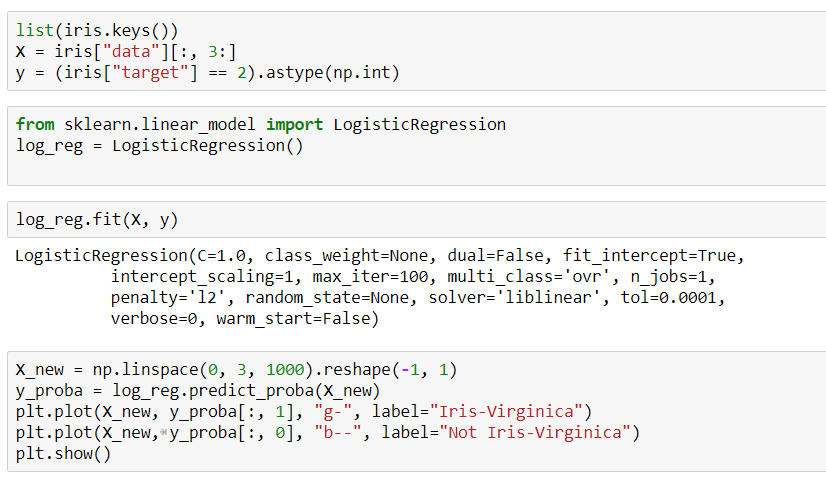
**(1)**

We are basically going to Solve the binary classification problem. By first load the training data and then inspecting the data

Imported all the libraries needed for the project:



Ran the code available in the text book:



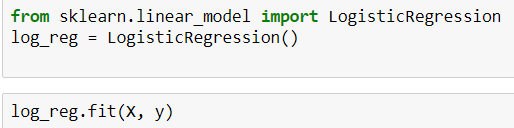
Basically, we are using the logistic regression here. It will act as classifier. Initially we imported the iris data set and stored it in X. Possible with the following commands



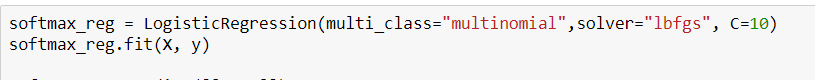
And



We will now use logistic regression with the following commands



X and y will be input.



We are using lbfgs solver and setting multi\_class parameter. To train the data we are going to use more than 2 classes.

**What is lbfgs solver?**

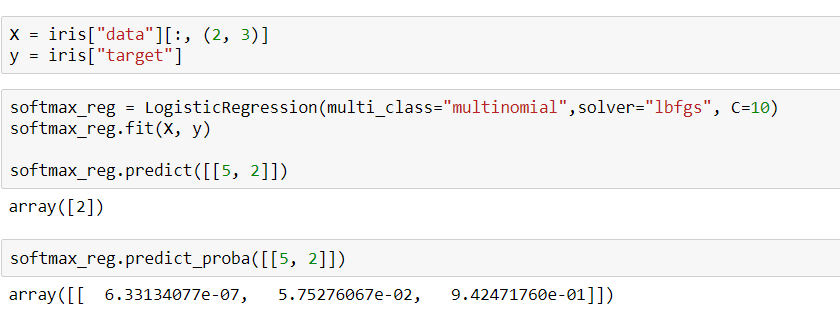
It’s limited memory BFGS optimization algorithm. If you are using this for a huge amount of data; classification problem is taken care by using optimization methods to save on memory

We have build a model. Now let’s test it. Say you find an iris and its 5 cm long and 2 cm wide petals. This model can predict if its Virginica with 94.2% probability.



Now, let’s go and **select one of the three class as positive**

Through done the following code; earlier I had explained how the code works.



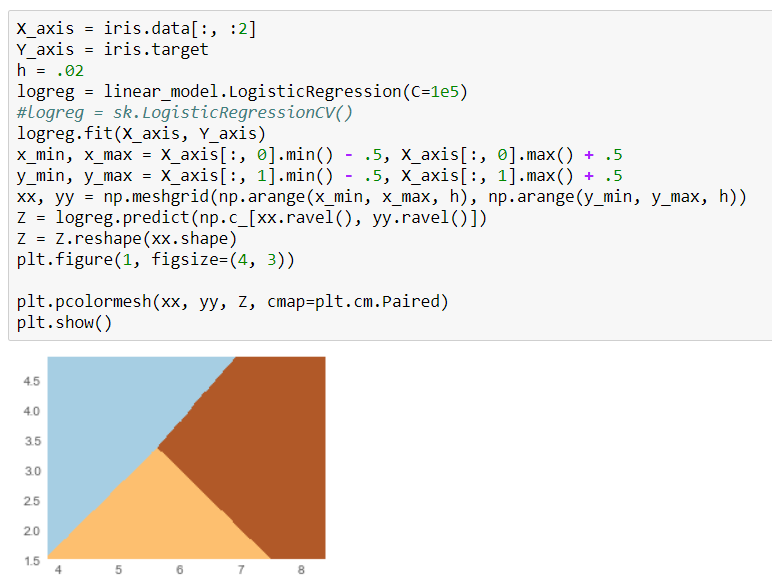
Coming to 2nd part of the Question

**(2)**

# Basically, we need to implement Logistic Regression 3-Class Classifier in Scikit-learn.

Which is done by the following code:

The 3 classes are depicted in the diagram below.



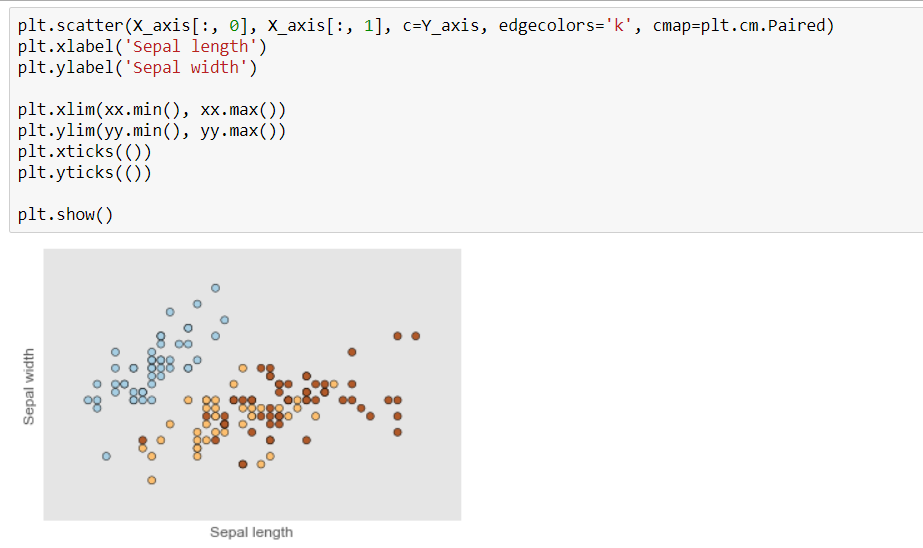
logreg = [linear\_model.LogisticRegression](http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html#sklearn.linear_model.LogisticRegression)(C=1e5)

or

logreg = sk.LogisticRegressionCV()

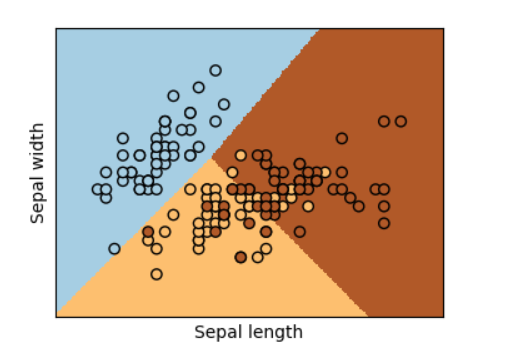
Both should work just fine. H is step size in above code. Pcolormesh is creating the colored mesh in the above diagram.

The following code will depict the Sepal length and Sepal width:



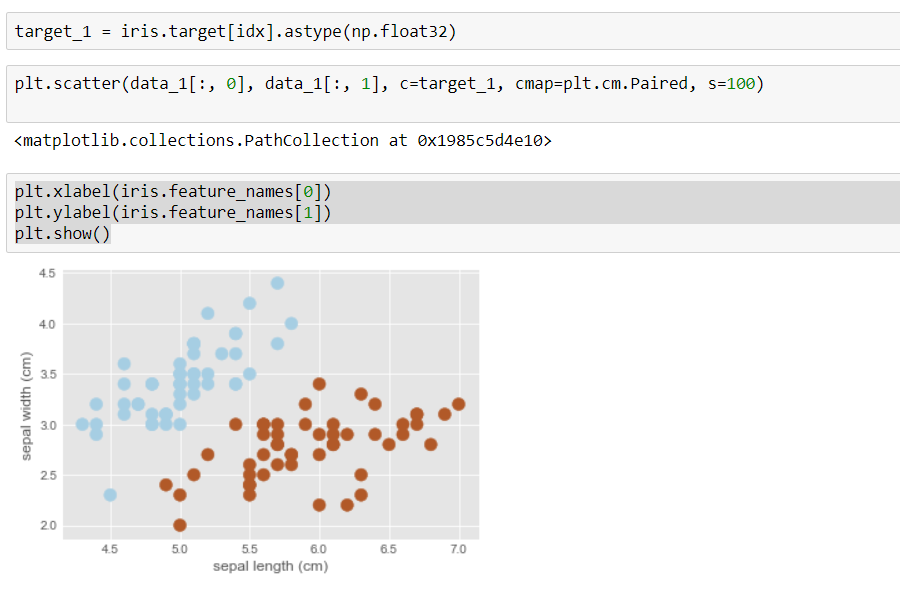
The data-points are colored according to their labels. Logistic-regression classifiers decision boundaries on the [iris](https://en.wikipedia.org/wiki/Iris_flower_data_set) dataset.

**Display a decision boundary diagram using the two other dimensions not shown in Figure 4-25. Is depicted here.**



The above implementation can be done in a unique/different way if needed





Reference :

[http://scikit-learn.org/stable/auto\_examples/linear\_model/plot\_iris\_logistic.html [1](http://scikit-learn.org/stable/auto_examples/linear_model/plot_iris_logistic.html%20%5b1)]

Hands on Machine learning with scikit learn & tensorflow Author: Aurelien Geron [2]

Source Code

