Hyperion Development

Task: Machine Learning II

www.hyperiondev.com

Introduction

Welcome to The Second Machine Learning Task!

This task is designed to give you a better understanding of the scikit-learn library, which we will be using for the remainder of the machine learning tasks.

You don't have to take our courses alone! This course has been designed to be taken with an online mentor that marks your submitted code by hand and supports you to achieve your career goals on a daily basis.

To access this mentor support simply navigate to www.hyperiondev.com/support.

A NOTE FROM THE HYPERION TEAM

If this is your first venture into machine learning, you may be wondering why learn/use scikit-learn? Well, for starters, scikit-learn is perhaps the most popular machine learning library available for the Python programming language. This is because scikit-learn provides simple and efficient tools for data mining and data analysis. It is also an open source library, which means you have complete license to edit, modify, and/or redistribute any code contained within the module - similarly to NumPy, Matplotlib, and SciPy. Thus, you can use the package as is or alternatively, you can tailor the package to suit your needs. Furthermore, if you plan to pursue a career in machine learning, it is worth learning scikit-learn to start familiarizing yourself with machine learning models (since scikit-learn encompasses some of the most popular and widely used machine learning algorithms) you can apply to different machine learning problems.

The Hyperion Team

How to install SciPy?

The "packages" folder contained in the task 2 folder has all the packages you will need for the remaining Machine learning tasks. If you have not copied this folder yet, it to your "***C:\Python27***" directory where your python files were installed. If you want to download the latest versions of these packages, click on the links: NumPy, SciPy, Matplotlib and save them to the "packages" folder.

Installing SciPy via the internet

- If you have already installed **Microsoft Visual C++ Compiler for Python 2.7**, proceed to **Step 2** else double-click on the Microsoft Visual C++ Compiler for

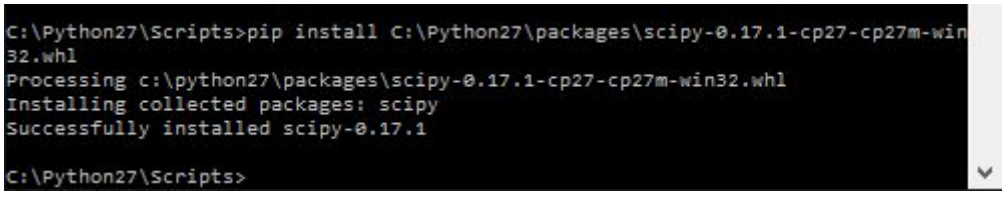
Python 2.7 (**VCForPython27.msi**) file to install it. If you have Visual Studio installed, you don't have to install this compiler as Visual Studio comes with it.

- Open "cmd" - command prompt by typing "cmd" on the windows search bar or press the windows key + r then type "cmd" and press Enter key.
- Once the command prompt has launched, type ***cd C:\Python27\Scripts*** and press Enter.
- The command for installing any package in python is `pip install package_name`. To install SciPy type "`pip install scipy`" and press Enter.

This is the easy way of installing packages. However, if you encounter problems or you don't have internet access, follow the steps below on how to install SciPy locally. You are going to need the "packages" folder that came with this task folder.

Installing SciPy locally

- If you have already installed **Microsoft Visual C++ Compiler for Python 2.7**, proceed to **Step 2** else double-click on the Microsoft Visual C++ Compiler for Python 2.7 (**VCForPython27.msi**) file to install it. If you have Visual Studio installed, you don't have to install this compiler as Visual Studio comes with it.
- Open "cmd" - command prompt by typing "cmd" on the windows search bar or press the windows key + r then type "cmd" and press Enter key.
- Once the command prompt has launched, type ***cd C:\Python27\Scripts*** and press Enter.
- Install SciPy by typing
  pip install C:\Python27\packages\scipy-0.17.1-cp27-cp27m-win32.whl and press Enter. You should get the same screen output as one below.



If you did not get any errors then you have successfully installed the SciPy package and you are ready to learn using SciPy . If you have problems completing the installation, contact your tutor to assist you.

Installing SciKit learn via the internet

- If you have already installed **Microsoft Visual C++ Compiler for Python 2.7**, proceed to **Step 2** else double-click on the Microsoft Visual C++ Compiler for

Python 2.7 (**VCForPython27.msi**) file to install it. If you have Visual Studio installed, you don't have to install this compiler as Visual Studio comes with it.

- The SciKit learn website has a webpage that covers installations across various operating systems. We will only cover installing SciKit for windows.

- Open "cmd" - command prompt by typing "cmd" on the windows search bar or press the windows key + r then type "cmd" and press Enter key.

- Once the command prompt has launched, type ***cd C:\Python27\Scripts*** and press Enter.

- The command for installing any package in python is `pip install package_name`. To install SciKit type "`pip install sklearn`" and press Enter.

Installing SciKit learn locally

- If you have already installed **Microsoft Visual C++ Compiler for Python 2.7**, proceed to **Step 2** else double-click on the Microsoft Visual C++ Compiler for Python 2.7 (**VCForPython27.msi**) file to install it. If you have Visual Studio installed, you don't have to install this compiler as Visual Studio comes with it.

- Open "cmd" - command prompt by typing "cmd" on the windows search bar or press the windows key + r then type "cmd" and press Enter key.

- Once the command prompt has launched, type ***cd C:\Python27\Scripts*** and press Enter.

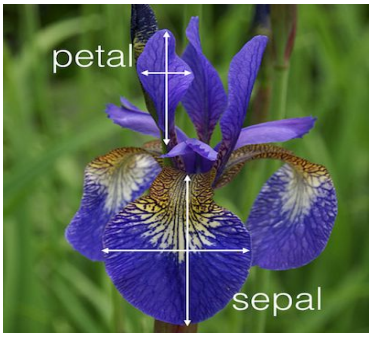
- Install SciKit learn by typing
pip install C:\Python27\packages\scikit_learn-0.17.1-cp27-cp27m-win32.whl and press Enter.

A NOTE FROM CEO
RIAZ MOOLA

Sorry to interrupt, but did you know that scikit-learn is an industry favourite and is used by popular websites such as Spotify, Inria, Evernote, and more. Spotify uses machine learning to give you suggestions to your favourite songs, whereas, Inria, a public research body dedicated to science and technology, uses it to support cutting-edge research for their computer vision and medical analyses. Evernote, on the other hand, uses scikit-learn to perform classifications and predictions.

SciKit learn Example using the iris dataset

Let's load the popular iris dataset in pattern recognition using scikit learn. Before we do that, here is a short description of the iris dataset. The iris dataset contains 3 classes of 50 instances each; a total of 150 observations of the iris flower specifying some measurements: sepal length, sepal width, petal length and petal width. Each class refers to the type of iris plant. The three classes are Setosa, Versicolour and Virginica. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other. The goal is to predict the class of the iris plant. Hence, it is said to be supervised learning.



Loading the dataset

There are two ways in which you can load a dataset from sklearn. If you are using a slow PC or laptop this might take a few seconds.

1. The first is to use sklearn module to import all datasets and then use the dot (.) notation to specify which dataset to load.

```
1. from sklearn import datasets
2. iris = datasets.load_iris()
```

2. Another way is to use sklearn.datasets module to import load_iris. Using this approach will omit the use of the dot (.) notation and just invoke the load_iris() function directly.

```
1. from sklearn.datasets import load_iris
2. iris = load_iris()
```

The 2nd option is considered convenient since you load only the dataset you will be working on rather than all sklearn datasets which you might not use. Sklearn uses the data type called "bunch" which is a container storing datasets and their attributes.

Iris dataset "bunch" attributes

These are the attributes contained in the iris dataset, namely; target_names, data, target, DESCR, and feature_names. Let's print out these attributes and the dataset object type.

```
1. print(type(iris))
2. print(iris['target_names'])
3. print(iris['data'])
4. print(iris['target'])
5. print(iris['DESCR'])
6. print(iris['feature_names'])
```

You can print out the dataset object with all its attributes and type by typing:

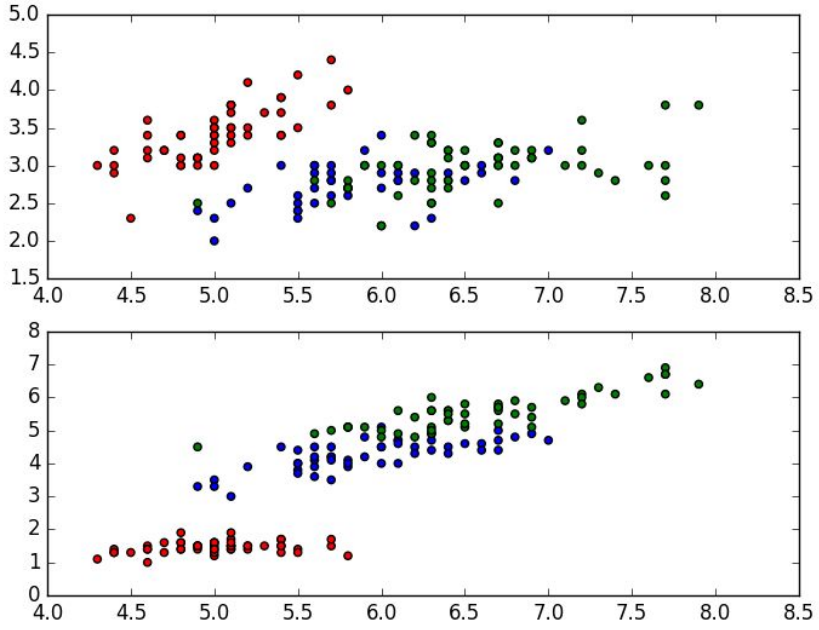
```
1. print(iris) #iris is the dataset object
```

Scatterplot of the iris dataset

Let's make a scatterplot of the iris dataset. The first thing to do is to import any packages we to be used i.e. matplotlib.pyplot, sklearn.datasets and numpy. In line 5 to 7, a figure and subplots are created. In line 9 to 11, an iris dataset is loaded to the "iris" object and then the iris object is used to extract the information in data and target attributes. In line 13, a dictionary is created to reference colors for the 3 classes. The reason why 0, 1 and 2 are used is that they represent the 3 classes, namely; Setosa, Versicolor and Virginica. Next, an array of colors is created so that the 3 classes are plotted in different colors.

Lastly, the scatterplots are plotted in their respective axes and shown. The first scatterplot uses column 0 and 1. The second scatterplot uses column 0 and 1.

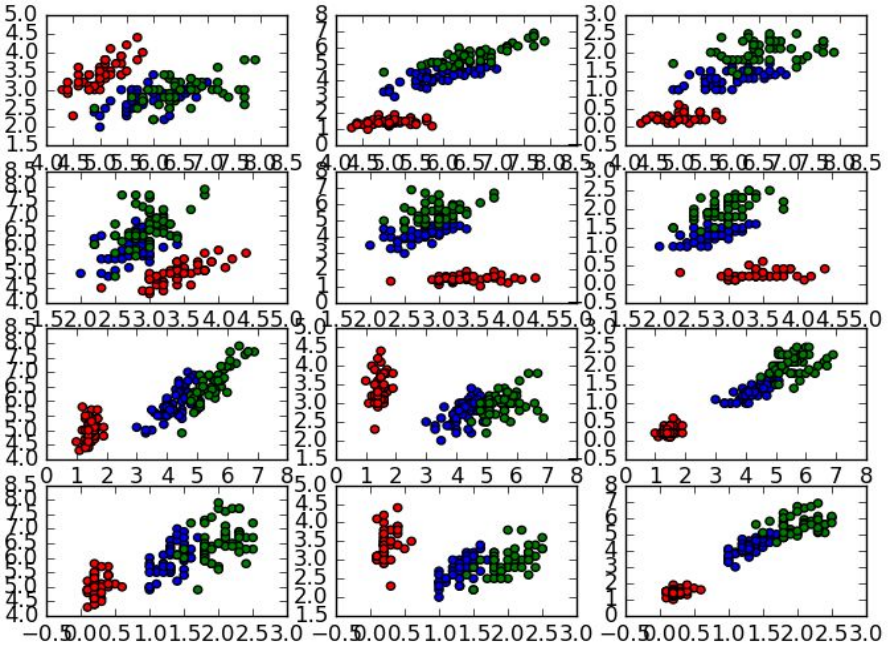
```
1.  import matplotlib.pyplot as plt
2.  from sklearn.datasets import load_iris
3.  import numpy as np
4.
5.  fig = plt.figure()
6.  ax1 = plt.subplot(2,1,1)
7.  ax2 = plt.subplot(2,1,2)
8.
9.  iris = load_iris()
10. data = np.array(iris['data'])
11. targets = np.array(iris['target'])
12.
13. cd = {0:'r',1:'b',2:"g"}
14. cols = np.array([cd[target] for target in targets])
15.
16. ax1.scatter(data[:,0], data[:,1], c=cols)
17. ax2.scatter(data[:,0], data[:,2], c=cols)
18. plt.show()
```



Compulsory Task

Follow these steps:

- Read Example.py
- Create a python file called IrisDatasetPlot.py.
- Write code to produce the following figure and save your figure.



- Hint: Note that there are 2 features being used for each subplot. i.e.The first 2 subplots came from our example of plotting the iris dataset. This is a permutation problem. From 4 features choose 2 which equals 12 choices in total. That is why there are 12 subplots in total.

SHARE YOUR THOUGHTS WITH US

Hyperion strives to provide internationally-excellent course content that helps you achieve your learning outcomes. Think the content of this task, or this course as a whole, can be improved or think we've done a good job? Click here to share your thoughts anonymously.