Neural Network and Learning Machines
Lab 02: Introduction to python – Part 2

Objectives

1. Setting up environment
2. Loading data
3. Exploring data
4. Visualizing data

Setting up environment

1. Install **python**
2. Install **an IDE**
3. Check that your system contains needed libraries
    a. Test by importing
    b.
```
import sys
import scipy
import numpy
import matplotlib
import pandas
import sklearn
```
4. If you face errors in importing libraries start downloading using the following commands (
```
conda install scipy
conda install pandas
conda install scikit-learn
conda install cycler
conda install -c conda-forge matplotlib
conda install -c conda-forge/label/broken matplotlib
conda install -c conda-forge/label/testing matplotlib
conda install -c conda-forge/label/rc matplotlib
```
5. Open **python IDE** and start a new project

Loading data

We are going to use iris flowers dataset. The dataset contains 150 observations of iris flowers. There are four columns of measurements of the flowers in centimeters. the length and the width of the sepals and petals. The fifth column is the species of the flower observed. All observed flowers belong to one of three species.

We are going to load the iris data from CSV file. We are using pandas to load the data. First import all modules, functions and objects we are going to use.

In :
```
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
dataset = pandas.read_csv(url, names=names)
```

Url can be replaced by the local path of the file. Note that we are specifying the names of each column when loading the data. This will help later in exploring the data.

Neural Network and Learning Machines
Lab 02: Introduction to python – Part 2

Exploring data

```
In :     print(dataset.head())
         # default n = 5
         # print(dataset.head(10))
```

```
Out :       sepal-length  sepal-width  petal-length  petal-width        class
        0           5.1          3.5           1.4          0.2  Iris-setosa
        1           4.9          3.0           1.4          0.2  Iris-setosa
        2           4.7          3.2           1.3          0.2  Iris-setosa
        3           4.6          3.1           1.5          0.2  Iris-setosa
        4           5.0          3.6           1.4          0.2  Iris-setosa
```

```
In : print(dataset.tail())
         # default n = 5
         # print(dataset.tail(10))
```

```
Out :       sepal-length  sepal-width  petal-length  petal-width           class
     145           6.7          3.0           5.2          2.3  Iris-virginica
     146           6.3          2.5           5.0          1.9  Iris-virginica
     147           6.5          3.0           5.2          2.0  Iris-virginica
     148           6.2          3.4           5.4          2.3  Iris-virginica
     149           5.9          3.0           5.1          1.8  Iris-virginica
```

```
In : print(dataset[4:6])
```

```
Out :       sepal-length  sepal-width  petal-length  petal-width        class
        4           5.0          3.6           1.4          0.2  Iris-setosa
        5           5.4          3.9           1.7          0.4  Iris-setosa
```

```
In : print(dataset.values)
```

```
Out :
        [[5.1 3.5 1.4 0.2 'Iris-setosa']
        [4.9 3.0 1.4 0.2 'Iris-setosa']
        [4.7 3.2 1.3 0.2 'Iris-setosa']
        ...
        [6.3 2.5 5.0 1.9 'Iris-virginica']
        [6.5 3.0 5.2 2.0 'Iris-virginica']
        [6.2 3.4 5.4 2.3 'Iris-virginica']
        [5.9 3.0 5.1 1.8 'Iris-virginica']]
```

```
In : print(dataset.columns)
```

```
Out :
        Index(['sepal-length', 'sepal-width', 'petal-length', 'petal-width',
'class'], dtype='object')
```

```
In : print(dataset.index)
```

```
Out :
        RangeIndex(start=0, stop=150, step=1)
```

```
In : print(dataset.T)
```

```
Out :
                            0            1            2            3            4  \
sepal-length              5.1          4.9          4.7          4.6            5
sepal-width              3.5            3          3.2          3.1          3.6
petal-length             1.4          1.4          1.3          1.5          1.4
petal-width              0.2          0.2          0.2          0.2          0.2
class            Iris-setosa  Iris-setosa  Iris-setosa  Iris-setosa  Iris-setosa
                         ...                  140            141            142  \
sepal-length             ...                  6.7            6.9            5.8
sepal-width              ...                  3.1            3.1            2.7
petal-length             ...                  5.6            5.1            5.1
petal-width              ...                  2.4            2.3            1.9
class                    ...       Iris-virginica  Iris-virginica  Iris-virginica
                         147            148            149
sepal-length             6.5            6.2            5.9
sepal-width                3            3.4              3
petal-length             5.2            5.4            5.1
petal-width                2            2.3            1.8
```

```
class           Iris-virginica  Iris-virginica  Iris-virginica
[5 rows x 150 columns]
```

In: `sorted_dataset = dataset.sort_index(axis = 0 , ascending = False);`
   `# , inplace = True);`
   `print(sorted_dataset.head())`

Out:
```
      sepal-length  sepal-width  petal-length  petal-width         class
149            5.9          3.0           5.1          1.8  Iris-virginica
148            6.2          3.4           5.4          2.3  Iris-virginica
147            6.5          3.0           5.2          2.0  Iris-virginica
146            6.3          2.5           5.0          1.9  Iris-virginica
145            6.7          3.0           5.2          2.3  Iris-virginica
```

In: `sorted_dataset = dataset.sort_values(by = 'sepal-length')`
   `print(sorted_dataset.head())`

Out:
```
    sepal-length  sepal-width  petal-length  petal-width        class
13           4.3          3.0           1.1          0.1  Iris-setosa
42           4.4          3.2           1.3          0.2  Iris-setosa
38           4.4          3.0           1.3          0.2  Iris-setosa
8            4.4          2.9           1.4          0.2  Iris-setosa
41           4.5          2.3           1.3          0.3  Iris-setosa
```

In: `print(dataset['sepal-width'])`

Out:
```
0      3.5
1      3.0
2      3.2
      ...
146    2.5
148    3.4
149    3.0
```

In: `print (dataset[dataset.get('sepal-width') > 4])`

Out:
```
    sepal-length  sepal-width  petal-length  petal-width        class
15           5.7          4.4           1.5          0.4  Iris-setosa
32           5.2          4.1           1.5          0.1  Iris-setosa
33           5.5          4.2           1.4          0.2  Iris-setosa
```

In: `print(dataset.shape)`

Out: `(150, 5)`

In: `print(dataset.groupby('class').size())`

Out:
```
class
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
dtype: int64
```

In: `print(dataset.mean())`

Out:
```
sepal-length    5.843333
sepal-width     3.054000
petal-length    3.758667
petal-width     1.198667
dtype: float64
```

In: `print(dataset.mean(1))`

Out:
```
0      2.550
1      2.375
      ...
148    4.325
149    3.950
Length: 150, dtype: float64
```
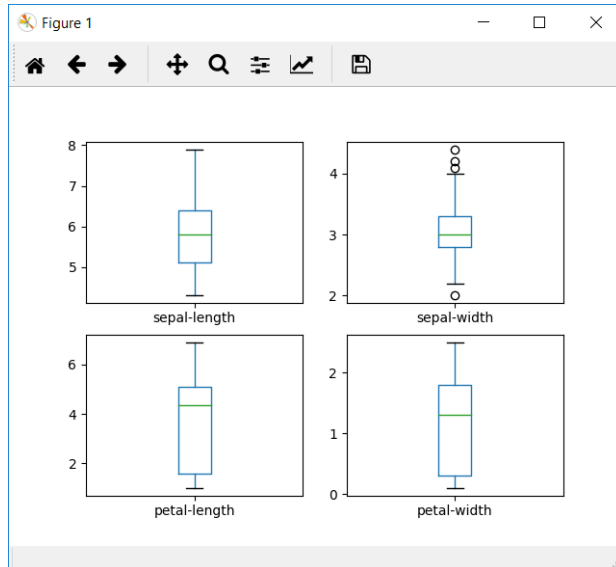
Neural Network and Learning Machines
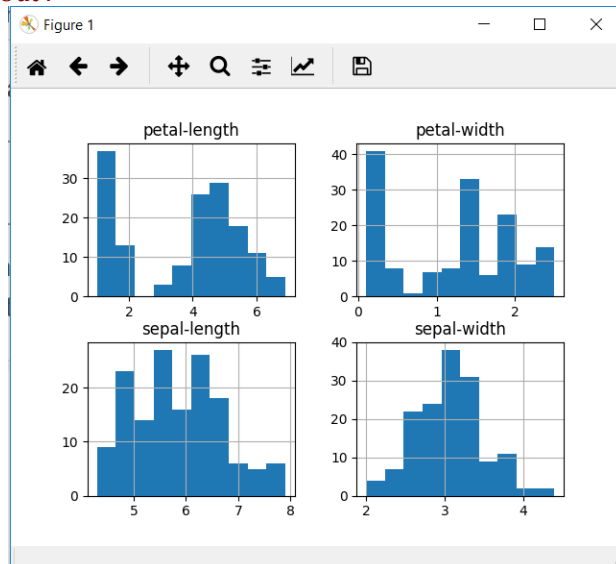**Lab 02**: Introduction to python – Part 2

Visualizing Data

In :
```
# add import: import matplotlib.pyplot as plt
dataset.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()
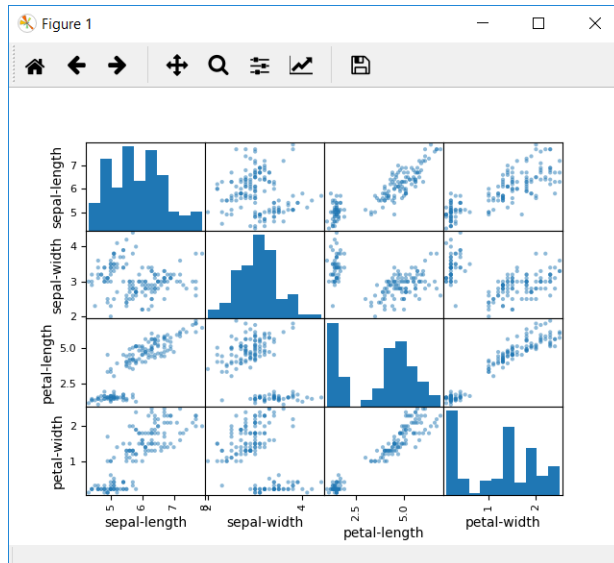```
Out :



In :
```
dataset.hist()
plt.show()
```
Out :



In :
```
#add import: from pandas.plotting import scatter_matrix
scatter_matrix(dataset)
plt.show()
```

# Neural Network and Learning Machines
**Lab 02**: Introduction to python – Part 2



## References

- https://machinelearningmastery.com/machine-learning-in-python-step-by-step/
- https://pandas.pydata.org/pandas-docs/stable/10min.html
-