



Converting SKLEARN utils.Bunch To pd.DataFrame

by Jesse P. Gutierrez Jr, 02/23/2019

When using sklearn embedded datasets they "type" of the set will be .utils.Bunch, this will make working with the set difficult. So it is a good idea to convert it to a pd.DataFrame so you can use Pandas.

```
In [1]: from sklearn import datasets  
import pandas as pd
```

```
In [2]: iris = datasets.load_iris()
```

```
In [18]: type(iris)
```

```
Out[18]: sklearn.utils.Bunch
```

```
Out[8]: {'data': array([[5.1, 3.5, 1.4, 0.2],
                        [4.9, 3. , 1.4, 0.2],
                        [4.7, 3.2, 1.3, 0.2],
                        [4.6, 3.1, 1.5, 0.2],
                        [5. , 3.6, 1.4, 0.2],
                        [5.4, 3.9, 1.7, 0.4],
                        [4.6, 3.4, 1.4, 0.3],
                        [5. , 3.4, 1.5, 0.2],
                        [4.4, 2.9, 1.4, 0.2],
                        [4.9, 3.1, 1.5, 0.1],
                        [5.4, 3.7, 1.5, 0.2],
                        [4.8, 3.4, 1.6, 0.2],
```

The type(iris) tell us that the data is a sklearn.utils.Bunch and will be presented in the above image.

```
In [19]: df = pd.DataFrame(data=iris.data,
                           columns=iris.feature_names)
```

```
In [20]: df.head()
```

```
Out[20]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

```
In [21]: df.tail()
```

```
Out[21]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

In [25]: df.info()

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 150 entries, 0 to 149  
Data columns (total 4 columns):  
sepal length (cm)    150 non-null float64  
sepal width (cm)     150 non-null float64  
petal length (cm)    150 non-null float64  
petal width (cm)     150 non-null float64  
dtypes: float64(4)  
memory usage: 4.8 KB
```