baseline

August 14, 2017

```
In [2]: import sqlite3
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
        from pylab import rcParams
        rcParams['figure.figsize'] = 12, 7
```

```
In [3]: color_scheme = dict(setosa='#902BFC', versicolor='#FF29FB', virginica='#2990FF')
        #color_scheme = ('#902BFC', '#FF29FB', '#2990FF')
```

```
In [4]: conn = sqlite3.connect('database.sqlite')
        cursor = conn.cursor()
        cursor.execute("SELECT name FROM sqlite_master WHERE type='table';")
        print(cursor.fetchall())
```

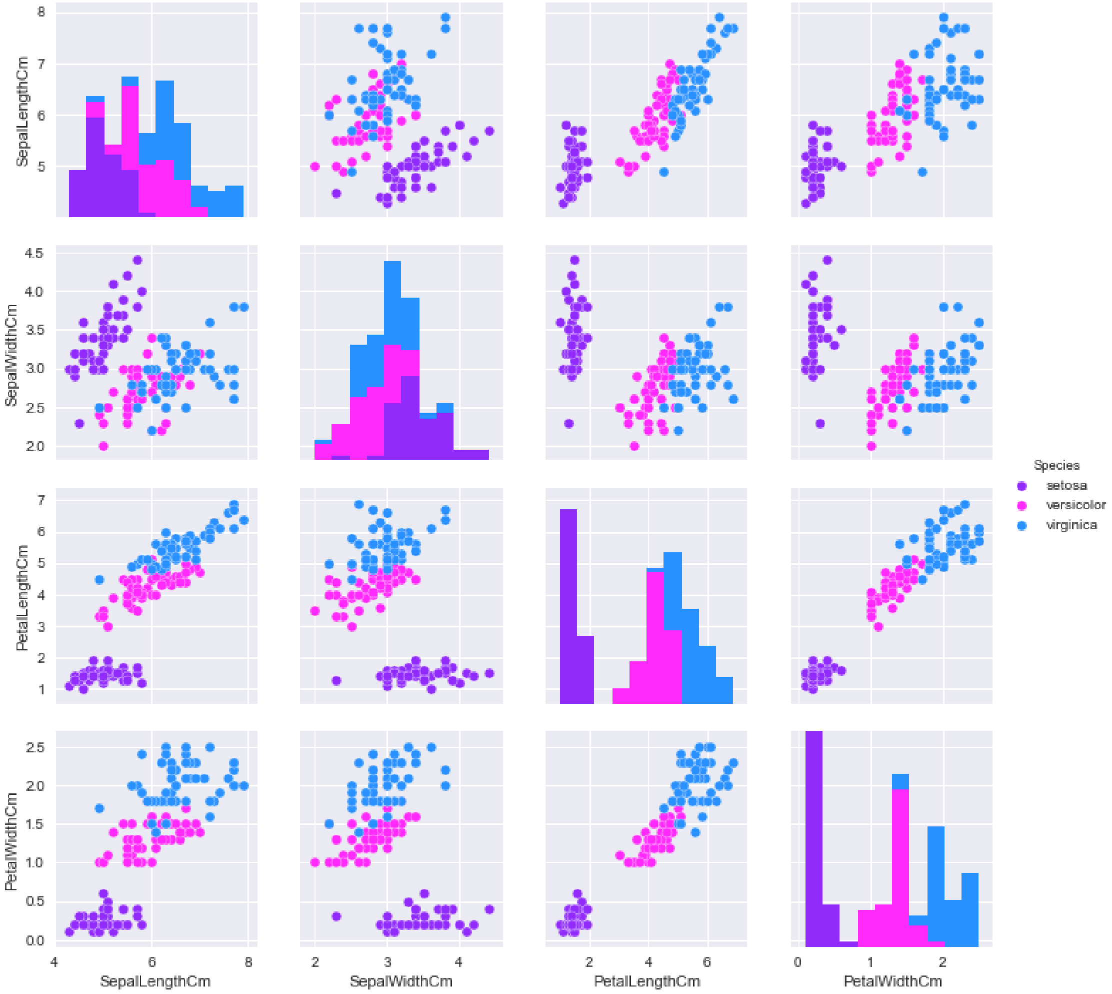
```
[('Iris',)]
```

```
In [5]: sql = "SELECT * FROM Iris"
        iris_df = pd.read_sql(sql, conn)
        iris_df.Species = iris_df.Species.str.replace('Iris-','')
        iris_df.head()
```

```
Out[5]:    Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm Species
        0   1            5.1           3.5            1.4           0.2  setosa
        1   2            4.9           3.0            1.4           0.2  setosa
        2   3            4.7           3.2            1.3           0.2  setosa
        3   4            4.6           3.1            1.5           0.2  setosa
        4   5            5.0           3.6            1.4           0.2  setosa
```

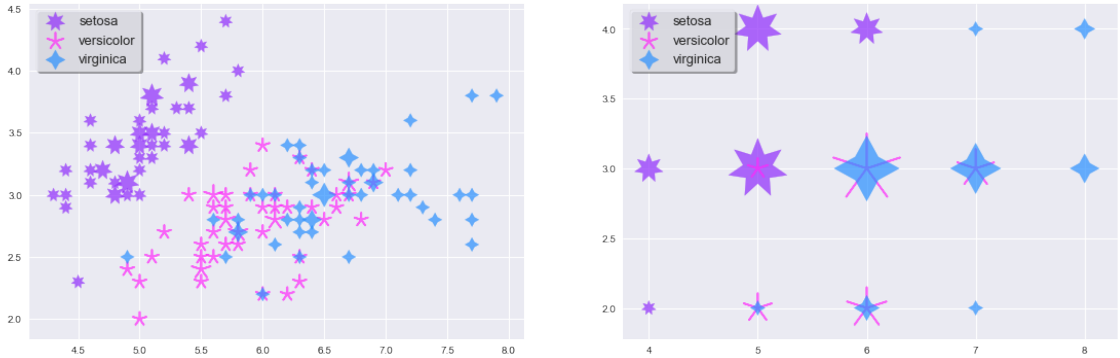
```
In [216]: sns.pairplot(iris_df[iris_df.columns[[1,2,3,4,5]]], hue='Species',
                       palette = color_scheme)
```

```
Out[216]: <seaborn.axisgrid.PairGrid at 0x1fd12854fd0>
```



```
In [13]: ax0_df = iris_df.groupby(['Species','SepalLengthCm','SepalWidthCm']).size().
         reset_index(name='cnt')

         setosa = ax0_df[ax0_df.Species == 'setosa']
         versicolor = ax0_df[ax0_df.Species == 'versicolor']
         virginica = ax0_df[ax0_df.Species == 'virginica']

         fig, axes = plt.subplots(1,2, figsize=(19,6))
         # First plot
         marker_size = 180
         l1 = axes[0].scatter(x=setosa.SepalLengthCm, y=setosa.SepalWidthCm, c ='#902BFC', s=se
         l2 = axes[0].scatter(x=versicolor.SepalLengthCm, y = versicolor.SepalWidthCm, c ='#FF2
         l3 = axes[0].scatter(x=virginica.SepalLengthCm, y = virginica.SepalWidthCm, c ='#2990
         axes[0].legend([l1,l2,l3], ['setosa','versicolor','virginica'], loc=2, fontsize=13, f

         # Second plot - Round dimensions
```

```
ax0_df.SepalLengthCm = ax0_df.SepalLengthCm.round()
ax0_df.SepalWidthCm = ax0_df.SepalWidthCm.round()
ax1_df = ax0_df.groupby(['Species','SepalLengthCm','SepalWidthCm']).size().reset_inde
setosa = ax1_df[ax1_df.Species == 'setosa']
versicolor = ax1_df[ax1_df.Species == 'versicolor']
virginica = ax1_df[ax1_df.Species == 'virginica']
leg = axes[1].legend([l1,l2,l3], ['setosa','versicolor','virginica'], loc=2, fontsize
marker_size = 200

l1 = axes[1].scatter(x=setosa.SepalLengthCm, y=setosa.SepalWidthCm, c ='#902BFC', s=se
l2 = axes[1].scatter(x=versicolor.SepalLengthCm, y = versicolor.SepalWidthCm, c ='#FF
l3 = axes[1].scatter(x=virginica.SepalLengthCm, y = virginica.SepalWidthCm, c ='#2990
```



```
In [23]: from sklearn.datasets import load_iris
         from sklearn import tree
         from sklearn.externals.six import StringIO
         import numpy as np
         import pydotplus
         from IPython.display import Image

         iris = load_iris()
         print (iris.feature_names)
         print (iris.target_names)
         print(iris.data[0])
         print(iris.target[0])

         for i, item in enumerate(iris.target):
             print("Example %d: label %s, features %s"%(i, item, iris.data[i]))
             if i == 5: break
```

```
['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']
['setosa' 'versicolor' 'virginica']
[ 5.1  3.5  1.4  0.2]
0
```

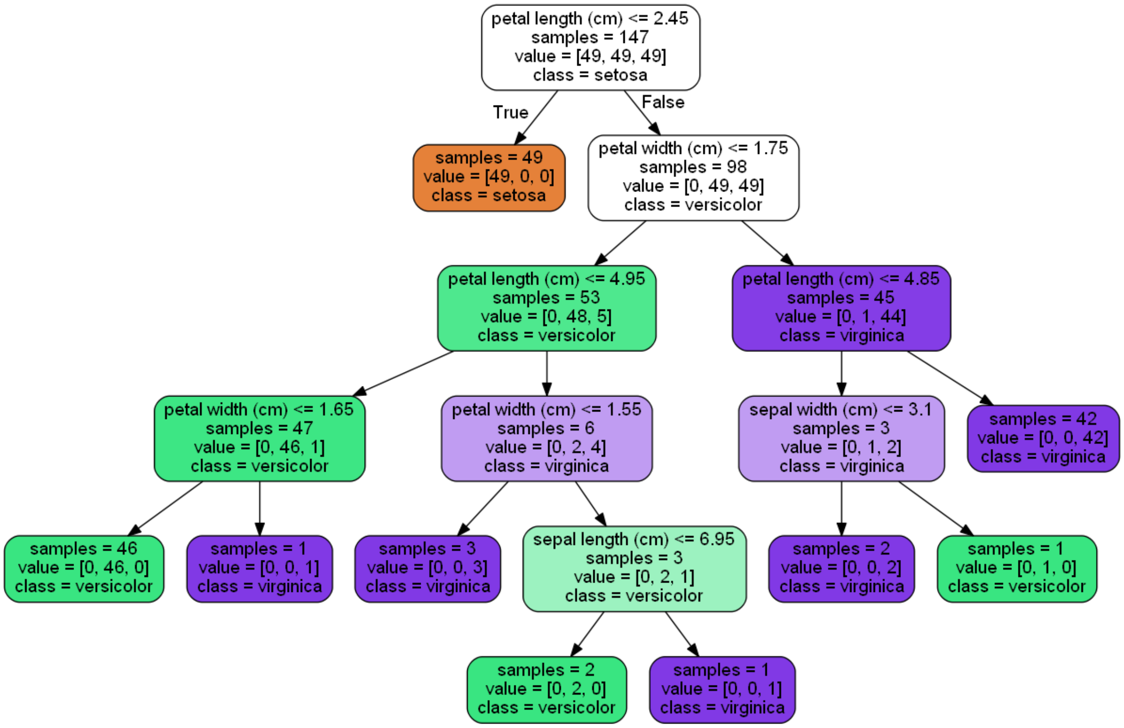
```
Example 0: label 0, features [ 5.1  3.5  1.4  0.2]
Example 1: label 0, features [ 4.9  3.   1.4  0.2]
Example 2: label 0, features [ 4.7  3.2  1.3  0.2]
Example 3: label 0, features [ 4.6  3.1  1.5  0.2]
Example 4: label 0, features [ 5.   3.6  1.4  0.2]
Example 5: label 0, features [ 5.4  3.9  1.7  0.4]
```

```
In [16]: test_idx = [0,50,100]

         #training data
         train_target = np.delete(iris.target, test_idx)
         train_data = np.delete(iris.data, test_idx, axis = 0)

         test_target = iris.target[test_idx]
         test_data = iris.data[test_idx]

         clf = tree.DecisionTreeClassifier()
         clf.fit(train_data, train_target)

         print(test_target)
         print(clf.predict(test_data))
```

```
[0 1 2]
[0 1 2]
```

```
In [24]: dot_data = StringIO()
         tree.export_graphviz(clf,
                              out_file = dot_data,
                              feature_names = iris.feature_names,
                              class_names = iris.target_names,
                              filled = True, rounded = True, impurity=False
                             )

         graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
         #graph.write_pdf('iris.pdf')
         Image(graph.create_png())
         #help(graph)
```

```
Out[24]:
```


petal length (cm) <= 2.45
samples = 147
value = [49, 49, 49]
class = setosa
True
False
samples = 49
value = [49, 0, 0]
class = setosa
petal width (cm) <= 1.75
samples = 98
value = [0, 49, 49]
class = versicolor
petal length (cm) <= 4.95
samples = 53
value = [0, 48, 5]
class = versicolor
petal length (cm) <= 4.85
samples = 45
value = [0, 1, 44]
class = virginica
petal width (cm) <= 1.65
samples = 47
value = [0, 46, 1]
class = versicolor
petal width (cm) <= 1.55
samples = 6
value = [0, 2, 4]
class = virginica
sepal width (cm) <= 3.1
samples = 3
value = [0, 1, 2]
class = virginica
samples = 42
value = [0, 0, 42]
class = virginica
samples = 46
value = [0, 46, 0]
class = versicolor
samples = 1
value = [0, 0, 1]
class = virginica
samples = 3
value = [0, 0, 3]
class = virginica
sepal length (cm) <= 6.95
samples = 3
value = [0, 2, 1]
class = versicolor
samples = 2
value = [0, 0, 2]
class = virginica
samples = 1
value = [0, 1, 0]
class = versicolor
samples = 2
value = [0, 2, 0]
class = versicolor
samples = 1
value = [0, 0, 1]
class = virginica