

```
1  """
2  Example 4.1 Introduction to the IRIS data set
3  @author: austin_downey
4  """
5
6  import IPython as IP
7  IP.get_ipython().run_line_magic('reset', '-sf')
8
9  import matplotlib.pyplot as plt
10 import sklearn as sk
11
12 cc = plt.rcParams['axes.prop_cycle'].by_key()['color']
13 plt.close('all')
14
15
16 %% Load your data
17
18 # We will use the Iris data set. This dataset was created by biologist Ronald
19 # Fisher in his 1936 paper "The use of multiple measurements in taxonomic
20 # problems" as an example of linear discriminant analysis
21
22 iris = sk.datasets.load_iris()
23
24 # for simplicity, extract some of the data sets
25 X = iris['data'] # this contains the length of the pedals and sepals
26 Y = iris['target'] # contains what type of flower it is
27 Y_names = iris['target_names'] # contains the name that aligns with the type of the
28 flower
29 feature_names = iris['feature_names'] # the names of the features
30
31 # plot the Sepal data
32 plt.figure(figsize=(6.5,3))
33 plt.subplot(121)
34 plt.grid(True)
35 plt.scatter(X[Y==0,0],X[Y==0,1],marker='o')
36 plt.scatter(X[Y==1,0],X[Y==1,1],marker='s')
37 plt.scatter(X[Y==2,0],X[Y==2,1],marker='d')
38 plt.xlabel(feature_names[0])
39 plt.ylabel(feature_names[1])
40
41 plt.subplot(122)
42 plt.grid(True)
43 plt.scatter(X[Y==0,2],X[Y==0,3],marker='o',label=Y_names[0])
44 plt.scatter(X[Y==1,2],X[Y==1,3],marker='s',label=Y_names[1])
45 plt.scatter(X[Y==2,2],X[Y==2,3],marker='d',label=Y_names[2])
46 plt.xlabel(feature_names[2])
47 plt.ylabel(feature_names[3])
48 plt.legend(framealpha=1)
49 plt.tight_layout()
50
```