Playing arund with Iris

We will use Iris in class to practice some attribute transformations and computing similarities.

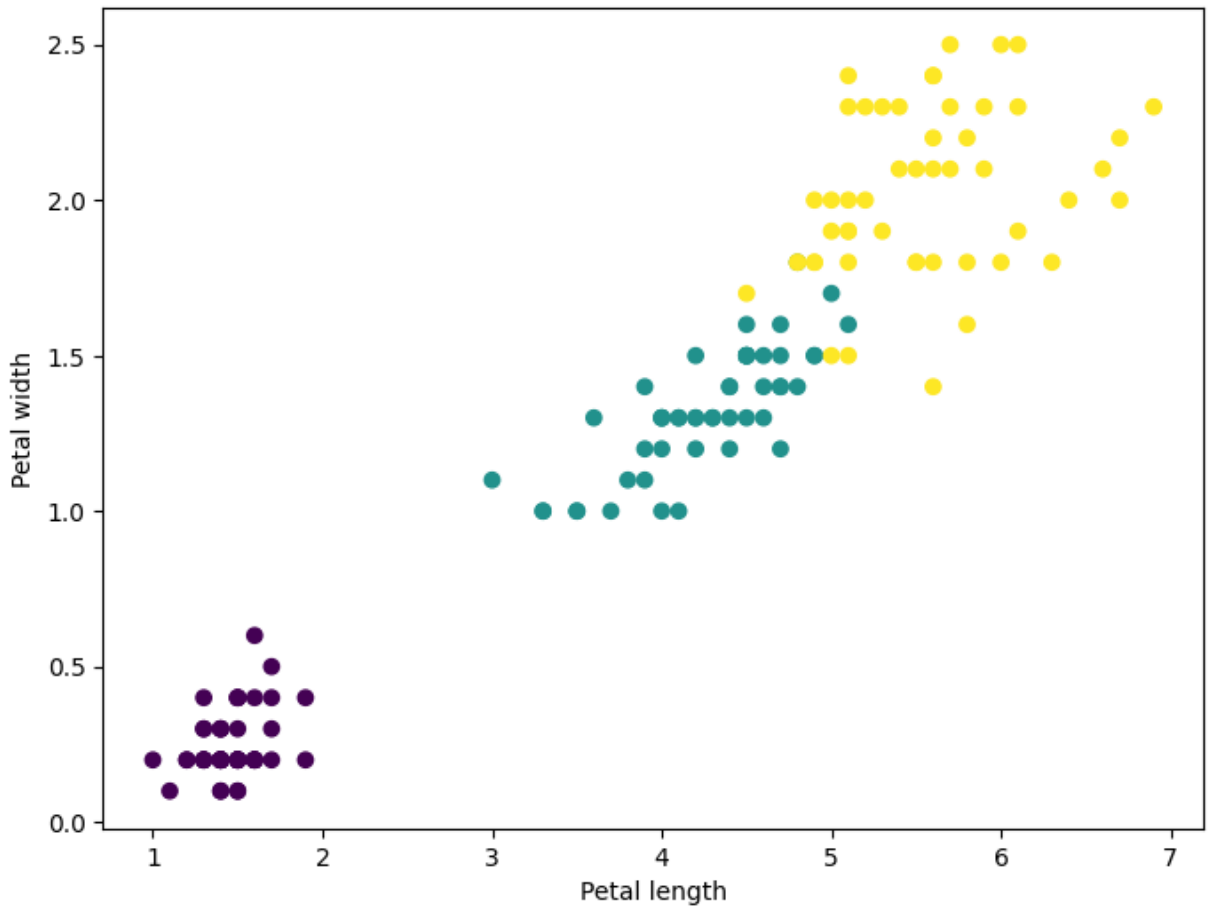
In [1]:
```
import matplotlib.pyplot as plt
from sklearn import datasets

# import some data to play with
iris = datasets.load_iris()
X = iris.data[:, 2:4]  # we only take petal length and petal width.
Y = iris.target

plt.figure(2, figsize=(8, 6))
plt.clf()

# Plot the training points
plt.scatter(X[:, 0], X[:, 1], c=Y)
plt.xlabel('Petal length')
plt.ylabel('Petal width')

plt.show()
```



In [26]:
```
import numpy as np
A = iris.data
print(A)
a = A[0,:] #Let's look at the first row
b = A[-1,:] #this is how to look at the last row
print(a,b)
iris.target
```

```
[[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]
 [4.8 3.  1.4 0.1]
 [4.3 3.  1.1 0.1]
 [5.8 4.  1.2 0.2]
 [5.7 4.4 1.5 0.4]
 [5.4 3.9 1.3 0.4]
 [5.1 3.5 1.4 0.3]
 [5.7 3.8 1.7 0.3]
```

In [4]:
```
c = np.log(a) # we can take the log values of a vector
print(a)
print(c)
```

```
[5.1 3.5 1.4 0.2]
[ 1.62924054  1.25276297  0.33647224 -1.60943791]
```

In [5]:
```
d = np.abs(c) # we can also take the ablsolute values
print(c)
print(d)
```

```
[ 1.62924054  1.25276297  0.33647224 -1.60943791]
[1.62924054 1.25276297 0.33647224 1.60943791]
```

In [6]:
```
for i in range(A.shape[1]): #shape gives the number of elements in array A for di
    print(np.min(A[:,i]), np.max(A[:,i])) #Let's print the min and max for each c
```

```
4.3 7.9
2.0 4.4
1.0 6.9
0.1 2.5
```

In [8]:
```
c = A[:,0] #Let's look at the first attribute (column)
```

TO DO: Use numpy to compute the mean, standard deviation, and z-score normalization for the first attribute. Print the mean and standard deviation. Then print the minimum, maximum, mean,

and standard deviation of the z-score normalized values.

In [30]:
```
import scipy.stats as stats
meanVal = np.mean(c) # Mean
sdValue = np.std(c) # Standard Deviation
print('Mean of c')
print(meanVal)
print('SD of c')
print(sdValue)
zscoreVal = stats.zscore(c) #zscore
print('Zscore of c')
print(zscoreVal)
#Min and Max
#for i in zscoreVal: #shape gives the number of elements in array A for dimensior
print(np.min(zscoreVal), np.max(zscoreVal))
print('Mean of zscoreVal')
print(np.mean(zscoreVal))
print('SD of zscoreVal')
print(np.std(zscoreVal))
```

```
Mean of c
5.843333333333334
SD of c
0.8253012917851409
Zscore of c
[-0.90068117 -1.14301691 -1.38535265 -1.50652052 -1.02184904 -0.53717756
 -1.50652052 -1.02184904 -1.74885626 -1.14301691 -0.53717756 -1.26418478
 -1.26418478 -1.87002413 -0.05250608 -0.17367395 -0.53717756 -0.90068117
 -0.17367395 -0.90068117 -0.53717756 -0.90068117 -1.50652052 -0.90068117
 -1.26418478 -1.02184904 -1.02184904 -0.7795133  -0.7795133  -1.38535265
 -1.26418478 -0.53717756 -0.7795133  -0.41600969 -1.14301691 -1.02184904
 -0.41600969 -1.14301691 -1.74885626 -0.90068117 -1.02184904 -1.62768839
 -1.74885626 -1.02184904 -0.90068117 -1.26418478 -0.90068117 -1.50652052
 -0.65834543 -1.02184904  1.40150837  0.67450115  1.2803405  -0.41600969
  0.79566902 -0.17367395  0.55333328 -1.14301691  0.91683689 -0.7795133
 -1.02184904  0.06866179  0.18982966  0.31099753 -0.29484182  1.03800476
 -0.29484182 -0.05250608  0.4321654  -0.29484182  0.06866179  0.31099753
  0.55333328  0.31099753  0.67450115  0.91683689  1.15917263  1.03800476
  0.18982966 -0.17367395 -0.41600969 -0.41600969 -0.05250608  0.18982966
 -0.53717756  0.18982966  1.03800476  0.55333328 -0.29484182 -0.41600969
 -0.41600969  0.31099753 -0.05250608 -1.02184904 -0.29484182 -0.17367395
 -0.17367395  0.4321654  -0.90068117 -0.17367395  0.55333328 -0.05250608
  1.52267624  0.55333328  0.79566902  2.12851559 -1.14301691  1.76501198
  1.03800476  1.64384411  0.79566902  0.67450115  1.15917263 -0.17367395
 -0.05250608  0.67450115  0.79566902  2.24968346  2.24968346  0.18982966
  1.2803405  -0.29484182  2.24968346  0.55333328  1.03800476  1.64384411
  0.4321654   0.31099753  0.67450115  1.64384411  1.88617985  2.4920192
  0.67450115  0.55333328  0.31099753  2.24968346  0.55333328  0.67450115
  0.18982966  1.2803405   1.03800476  1.2803405  -0.05250608  1.15917263
  1.03800476  1.03800476  0.55333328  0.79566902  0.4321654   0.06866179]
-1.870024133847019 2.4920192021244283
Mean of zscoreVal
-4.736951571734001e-16
SD of zscoreVal
1.0
```

In [ ]: