In [2]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn import datasets
from sklearn.metrics import accuracy_score
```

In [4]:

```python
iris = datasets.load_iris()
iris
```

Out[4]:

```
{'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],
       [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],
```

In [7]:

```python
dataS=np.array(iris.data)
dataS
```

Out[7]:

```
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],
       [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],
```

In [26]:

```python
covMat=pd.DataFrame(data=np.cov(dataS,rowvar=False),columns=iris.feature_names,index=iris.f
covMat
```

Out[26]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
sepal length (cm)	0.685694	-0.042434	1.274315	0.516271
sepal width (cm)	-0.042434	0.189979	-0.329656	-0.121639
petal length (cm)	1.274315	-0.329656	3.116278	1.295609
petal width (cm)	0.516271	-0.121639	1.295609	0.581006

In [31]:

```python
eignV, eignVe=np.linalg.eig(covMat)
```

In [32]:

```python
eignV
```

Out[32]:

```
array([4.22824171, 0.24267075, 0.0782095 , 0.02383509])
```

In [33]:

```python
eignVe
```

Out[33]:

```
array([[ 0.36138659, -0.65658877, -0.58202985,  0.31548719],
       [-0.08452251, -0.73016143,  0.59791083, -0.3197231 ],
       [ 0.85667061,  0.17337266,  0.07623608, -0.47983899],
       [ 0.3582892 ,  0.07548102,  0.54583143,  0.75365743]])
```

In [34]:

```python
pca2c=eignVe[:,:2]
```

In [35]:

```python
pca2c
```

Out[35]:

```
array([[ 0.36138659, -0.65658877],
       [-0.08452251, -0.73016143],
       [ 0.85667061,  0.17337266],
       [ 0.3582892 ,  0.07548102]])
```

In [36]:

```python
pca2cT=np.transpose(pca2c)
dataT=np.transpose(dataS)
newData=np.matmul(pca2cT,dataT)
newDataSet=np.transpose(newData)
newDataSet
```

Out[36]:

```
array([[ 2.81823951, -5.64634982],
       [ 2.78822345, -5.14995135],
       [ 2.61337456, -5.18200315],
       [ 2.75702228, -5.0086536 ],
       [ 2.7736486 , -5.65370709],
       [ 3.2215055 , -6.06828303],
       [ 2.68182738, -5.23749119],
       [ 2.87622016, -5.49033754],
       [ 2.6159824 , -4.74864082],
       [ 2.82960933, -5.21317833],
       [ 2.99541804, -5.97202148],
       [ 2.8896099 , -5.34168252],
       [ 2.71625587, -5.09184058],
       [ 2.27856139, -4.81555799],
       [ 2.85761474, -6.50571721],
       [ 3.1163261 , -6.66501491],
       [ 2.87883726, -6.13763209],
```

In [38]:

```python
color= np.array(['red','blue','green'])
plt.scatter(newDataSet[:,0],newDataSet[:,1],c=color[iris.target])
```

Out[38]:

```
<matplotlib.collections.PathCollection at 0x2b78253a310>
```
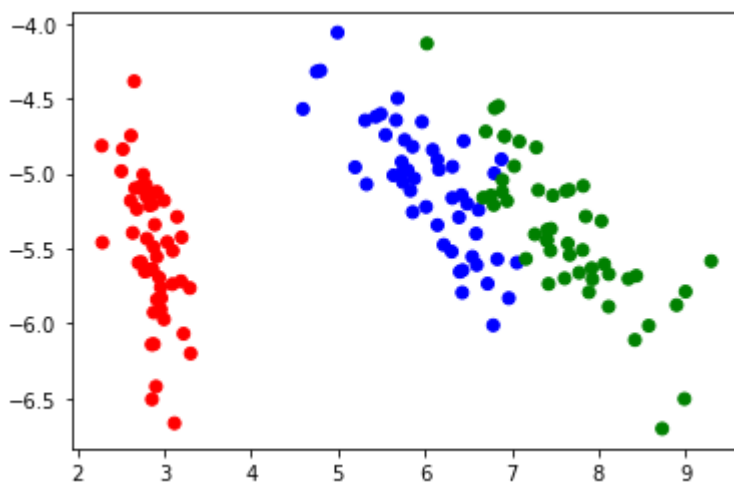
In [39]:

```python
a=KMeans(n_clusters=3)
a.fit(newDataSet)
```

Out[39]:

```
KMeans(n_clusters=3)
```

In [41]:
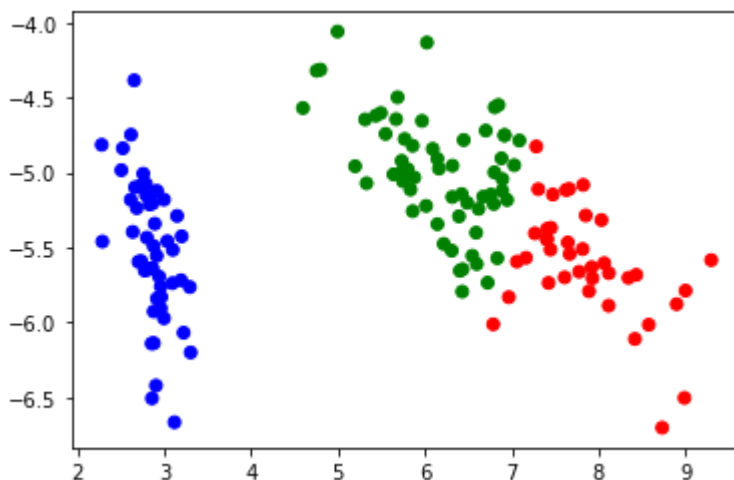
```python
a.labels_
```

Out[41]:

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 0, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 0, 0, 0, 0, 2, 0, 0, 0,
       0, 0, 0, 2, 2, 0, 0, 0, 0, 2, 0, 2, 0, 2, 0, 0, 2, 2, 0, 0, 0, 0,
       0, 2, 0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 2])
```

In [42]:

```python
color= np.array(['red','blue','green'])
plt.scatter(newDataSet[:,0],newDataSet[:,1],c=color[a.labels_])
```

Out[42]:

```
<matplotlib.collections.PathCollection at 0x2b782108a30>
```


In [43]:

```python
accuracy_score(iris.target,a.labels_)
```

Out[43]:

```
0.09333333333333334
```

In [ ]: