EXERCISE 10

1. Use iris flower dataset from sklearn library and try to form clusters of flowers using petal width and length features. Drop other two features for simplicity.
2. Figure out if any preprocessing such as scaling would help here
3. Draw elbow plot and from that figure out optimal value of k

In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
```

In [3]:

```
data = load_iris()
```

In [6]:

```
df = pd.DataFrame(data.data, columns=data.feature_names)
df.head()
```

Out[6]:

	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 [7]:

```
df.drop(['sepal length (cm)', 'sepal width (cm)'], axis=1, inplace=True)
```

In [8]:

```
df.head()
```

Out[8]:

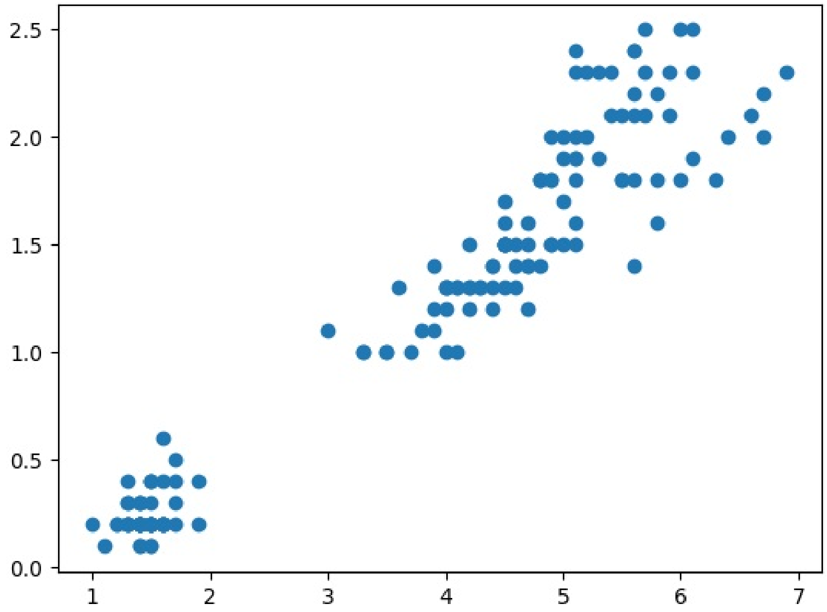
	petal length (cm)	petal width (cm)
0	1.4	0.2
1	1.4	0.2
2	1.3	0.2
3	1.5	0.2
4	1.4	0.2

In [12]:

```
plt.scatter(df['petal length (cm)'], df['petal width (cm)'])
```

Out[12]:

```
<matplotlib.collections.PathCollection at 0x2362e6672d0>
```



In [44]:

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

Out[44]:

```
▾ KMeans
KMeans(n_clusters=3)
```

In [45]:

```
predict =km.fit_predict(df[['petal length (cm)', 'petal width (cm)']])
predict
```

```
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```

Out[45]:

```
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, 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, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

In [46]:

```
df['cluster'] = predict
df.head()
```

Out[46]:

	petal length (cm)	petal width (cm)	cluster
0	0.067797	0.041667	1
1	0.067797	0.041667	1
2	0.050847	0.041667	1
3	0.084746	0.041667	1
4	0.067797	0.041667	1

In [47]:

```
df.shape
```

Out[47]:

```
(150, 3)
```

In [48]:

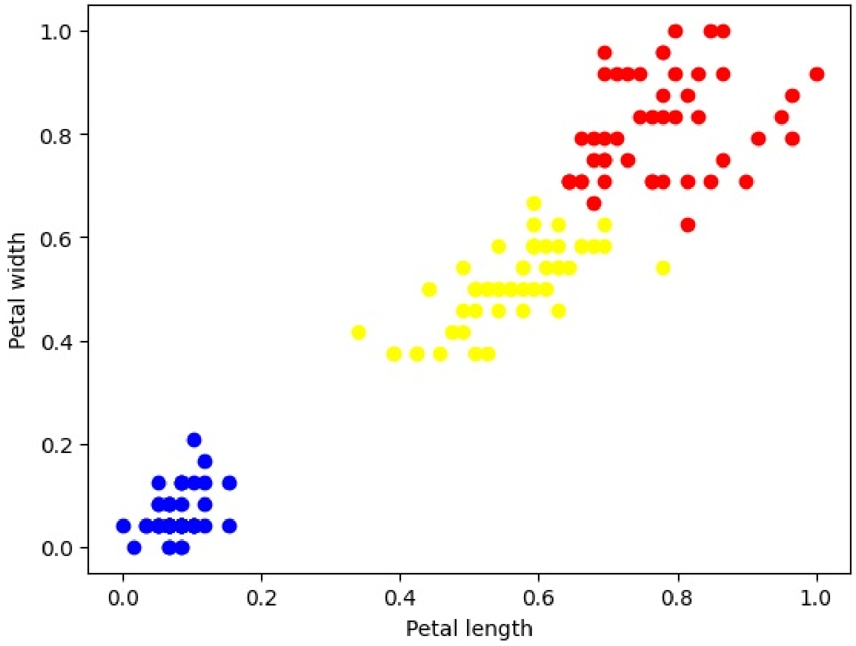
```
df0 = df[df.cluster == 0]
df1 = df[df.cluster == 1]
df2 = df[df.cluster == 2]
```

In [50]:

```
plt.scatter(df0['petal length (cm)'], df0['petal width (cm)'], color="red")
plt.scatter(df1['petal length (cm)'], df1['petal width (cm)'], color="blue")
plt.scatter(df2['petal length (cm)'], df2['petal width (cm)'], color="yellow")


plt.xlabel('Petal length')
plt.ylabel('Petal width')
```

Out[50]:

```
Text(0, 0.5, 'Petal width')
```



In [51]:

```
scaler = MinMaxScaler()
scaler.fit(df[['petal length (cm)']])
df['petal length (cm)'] = scaler.transform(df[['petal length (cm)']])
df
```

Out[51]:

	petal length (cm)	petal width (cm)	cluster
0	0.067797	0.041667	1
1	0.067797	0.041667	1
2	0.050847	0.041667	1
3	0.084746	0.041667	1
4	0.067797	0.041667	1
...	...	...	...
145	0.711864	0.916667	0
146	0.677966	0.750000	0
147	0.711864	0.791667	0
148	0.745763	0.916667	0
149	0.694915	0.708333	0

150 rows × 3 columns

In [52]:

```
scaler.fit(df[['petal width (cm)']])
df['petal width (cm)'] = scaler.transform(df[['petal width (cm)']])
df
```

Out[52]:

	petal length (cm)	petal width (cm)	cluster
0	0.067797	0.041667	1
1	0.067797	0.041667	1
2	0.050847	0.041667	1
3	0.084746	0.041667	1
4	0.067797	0.041667	1
...	...	...	...
145	0.711864	0.916667	0
146	0.677966	0.750000	0
147	0.711864	0.791667	0
148	0.745763	0.916667	0
149	0.694915	0.708333	0

150 rows × 3 columns

In [53]:

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

Out[53]:

```
▾ KMeans
KMeans(n_clusters=3)
```

In [54]:

```
predict = km.fit_predict(df[['petal length (cm)', 'petal width (cm)']])
predict
```

```
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```

Out[54]:

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

In [55]:

```
df['cluster'] = predict
df
```

Out[55]:

	petal length (cm)	petal width (cm)	cluster
0	0.067797	0.041667	0
1	0.067797	0.041667	0
2	0.050847	0.041667	0
3	0.084746	0.041667	0
4	0.067797	0.041667	0
...	...	...	...
145	0.711864	0.916667	1
146	0.677966	0.750000	1
147	0.711864	0.791667	1
148	0.745763	0.916667	1
149	0.694915	0.708333	1

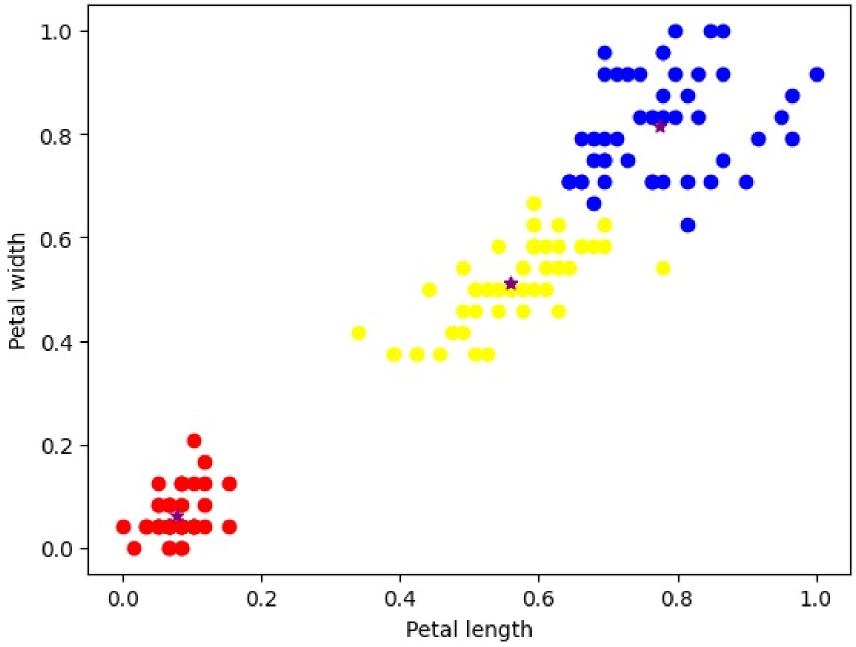
150 rows × 3 columns

In [56]:

```
df0 = df[df.cluster == 0]
df1 = df[df.cluster == 1]
df2 = df[df.cluster == 2]

plt.scatter(df0['petal length (cm)'], df0['petal width (cm)'], color="red")
plt.scatter(df1['petal length (cm)'], df1['petal width (cm)'], color="blue")
plt.scatter(df2['petal length (cm)'], df2['petal width (cm)'], color="yellow")

plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color='purple',marker='*',label='centroid')

plt.xlabel('Petal length')
plt.ylabel('Petal width')
```

Out[56]:

```
Text(0, 0.5, 'Petal width')
```



In [57]:

```
k_range = range(1, 10)
sse = []

for k in k_range:
    km = KMeans(n_clusters=k)
    km.fit(df[['petal length (cm)', 'petal width (cm)']])
    sse.append(km.inertia_)
```

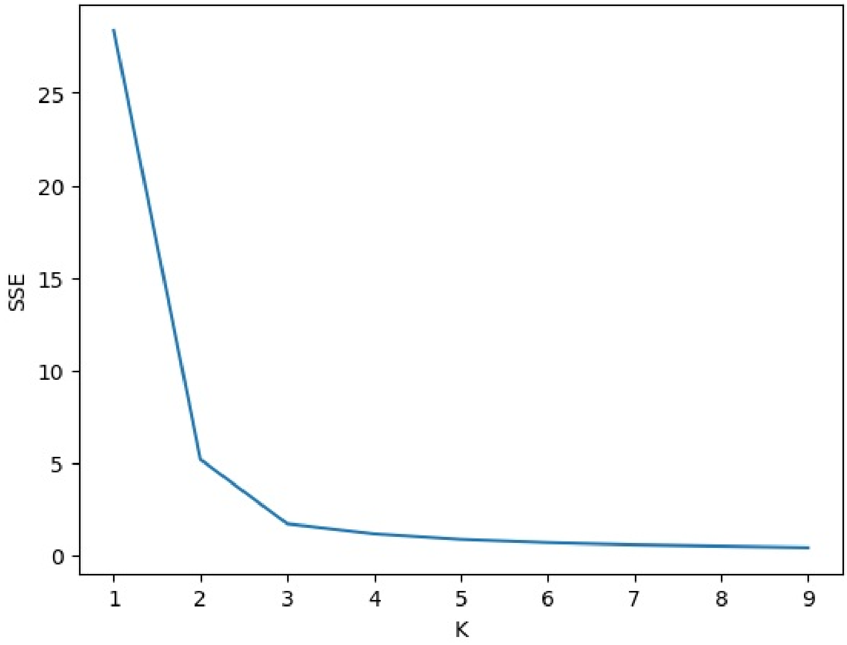
```
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureW
arning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly
to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```

In [58]:
```
sse
```

Out[58]:
```
[28.368353219727197,
 5.176463590044368,
 1.701874688192097,
 1.1588792731667124,
 0.8535683225340073,
 0.6801035704886194,
 0.5651273455703979,
 0.4858084553914369,
 0.4132667782512473]
```

In [59]:
```
plt.xlabel('K')
plt.ylabel('SSE')
plt.plot(k_range, sse)
```

Out[59]:
```
[<matplotlib.lines.Line2D at 0x23633ab6a90>]
```



In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js