In [2]:

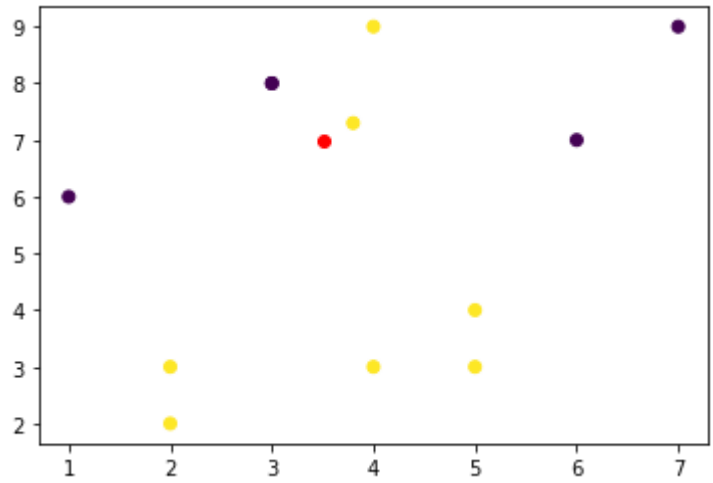
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
```

In [3]:

```
dff = pd.DataFrame({
    'x' : [1,2,3,4,5,6,7,5,2,3,4,3.8],
    'y' : [6,3,8,9,3,7,9,4,2,8,3,7.3],
    'z' : [0,1,0,1,1,0,0,1,1,0,1,1]
})
plt.scatter(dff['x'],dff['y'],c=dff['z'])
plt.scatter(3.5,7,color='r')
```

Out[3]:

```
<matplotlib.collections.PathCollection at 0x226427d3ac0>
```



In [8]:

```
def ed(x1,y1,x2,y2):
    return np.sqrt((x2-x1)**2+(y2-y1)**2)
ed(3.5,7,3,8)
```

Out[8]:

```
1.118033988749895
```

In [9]:

```
ed(3.5,7,3.8,7.3)
```

Out[9]:

```
0.4242640687119283
```

In [10]:

```
np.argmin([6.1,4.9])
```

Out[10]:

```
1
```

In [11]:

```
df = pd.read_csv('DATA/Social_Network_Ads.csv')
df
```

Out[11]:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
...	...	...	...	...	...
395	15691863	Female	46	41000	1
396	15706071	Male	51	23000	1
397	15654296	Female	50	20000	1
398	15755018	Male	36	33000	0
399	15594041	Female	49	36000	1

400 rows × 5 columns

In [12]:

```
df.drop(columns = 'User ID',inplace=True)
```

In [14]:

```
from sklearn.preprocessing import LabelEncoder,StandardScaler
lb = LabelEncoder()
sc = StandardScaler()
df['Gender'] = lb.fit_transform(df['Gender'])
df[['Age','EstimatedSalary']] = sc.fit_transform(df[['Age','EstimatedSalary']])
```

In [15]:

```
df
```

Out[15]:

	Gender	Age	EstimatedSalary	Purchased
0	1	-1.781797	-1.490046	0
1	1	-0.253587	-1.460681	0
2	0	-1.113206	-0.785290	0
3	0	-1.017692	-0.374182	0
4	1	-1.781797	0.183751	0
...	...	...	...	...
395	0	0.797057	-0.844019	1
396	1	1.274623	-1.372587	1
397	0	1.179110	-1.460681	1
398	1	-0.158074	-1.078938	0
399	0	1.083596	-0.990844	1

400 rows × 4 columns

In [16]:

```
ind = df.iloc[:,:3]
dep = df.iloc[:,-1]
dep
```

Out[16]:

```
0      0
1      0
2      0
3      0
4      0
      ..
395    1
396    1
397    1
398    0
399    1
Name: Purchased, Length: 400, dtype: int64
```

In [17]:

```
ind
```

Out[17]:

	Gender	Age	EstimatedSalary
0	1	-1.781797	-1.490046
1	1	-0.253587	-1.460681
2	0	-1.113206	-0.785290
3	0	-1.017692	-0.374182
4	1	-1.781797	0.183751
...	...	...	...
395	0	0.797057	-0.844019
396	1	1.274623	-1.372587
397	0	1.179110	-1.460681
398	1	-0.158074	-1.078938
399	0	1.083596	-0.990844

400 rows × 3 columns

In [19]:

```
xx = df.iloc
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(ind,dep,test_size=0.2,random_state=0)
```

In [20]:

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=2)
knn.fit(x_train,y_train)
```

Out[20]:

```
KNeighborsClassifier(n_neighbors=2)
```

In [21]:

```
testing = pd.DataFrame({
    'y':y_test,
    'y_hat': knn.predict(x_test)
})
testing.sample(8)
```

Out[21]:

	y	y_hat
361	1	1
1	0	0
68	0	0
176	0	0
45	0	0
354	1	0
318	1	1
363	0	0

In [22]:

```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,knn.predict(x_test))
```

Out[22]:

```
array([[55,  3],
       [ 3, 19]], dtype=int64)
```

In [23]:

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,knn.predict(x_test))
```

Out[23]:

```
0.925
```

In [ ]:

```

```