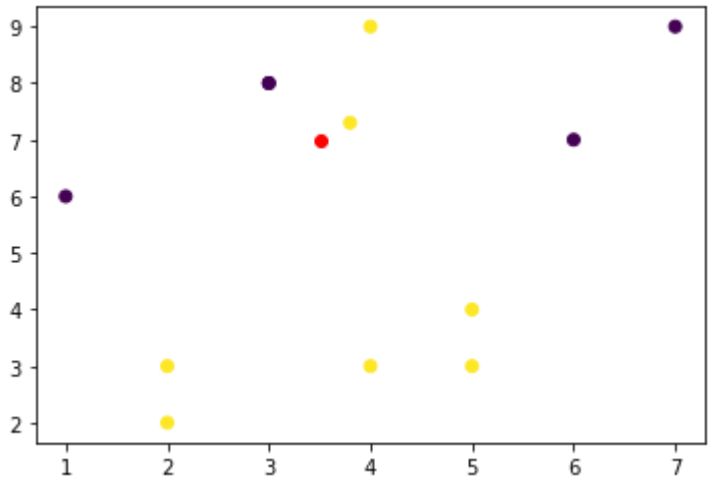
In [15]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from sklearn.datasets import make_classification
```

In [16]:
```
df1 = 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(df1['x'],df1['y'],c=df1['z'])
plt.scatter(3.5,7,color='r')
```

Out[16]:
```
<matplotlib.collections.PathCollection at 0x1684dc2f4f0>
```



In [17]:
```
#(3,8) (4,9)
def ed(x1,x2,z1,z2):
    return np.sqrt((x2-x1)**2 +(z2-z1)**2)
ed(3.5,7,3,8)
```

Out[17]:
```
6.103277807866851
```

In [18]:
```
ed(3.5,7,3.8,7.3)
```

Out[18]:
```
4.949747468305833
```

In [19]:
```
np.argmin([6.1,4.9])
```

Out[19]:
```
1
```

Customer Purchase

In [20]:
```
df = pd.read_csv('Social_Network_Ads.csv')
df
```

Out[20]:

	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 [21]:
```
df.drop(columns='User ID',inplace=True)
```

In [26]:
```
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 [27]:
```
df
```

Out[27]:

	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 [28]:
```
ind = df.iloc[:,:3]
dep = df.iloc[:,-1]
dep
```

Out[28]:
```
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 [29]:
```
ind
```

Out[29]:

	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 [30]:
```
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_
```

In [31]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=2)
knn.fit(x_train,y_train)
```

Out[31]:
```
KNeighborsClassifier(n_neighbors=2)
```

In [32]:
```
testing = pd.DataFrame({
    'y' : y_test,
    'y_hat' : knn.predict(x_test)
})
testing.sample(8)
```

Out[32]:

	y	y_hat
322	0	0
4	0	0
199	0	0
176	0	0
348	0	0
6	0	0
255	1	1
233	1	1

In [34]:
```
from sklearn.metrics import confusion_matrix,accuracy_score
confusion_matrix(y_test,knn.predict(x_test))
```

Out[34]:
```
array([[55,  3],
       [ 3, 19]], dtype=int64)
```

In [35]:
```
accuracy_score(y_test,testing['y_hat'])
```

Out[35]:
```
0.925
```

In [ ]:
```

```