In [9]:
```
import sklearn.datasets
data = sklearn.datasets.load_breast_cancer()
data

from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
import numpy as np
```

In [10]:
```
X = data.data
print(data.data.shape)
```

```
(569, 30)
```

In [11]:
```
y = data.target
print(data.target.shape)
```

```
(569,)
```

In [31]:
```
#Split into train and test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state =4) # 60% training and 40% test

# Create a k-NN classifier with 5 neighbors
knn = KNeighborsClassifier(n_neighbors=8)

# Fit the classifier to the training data
knn.fit(X_train, y_train)

#Predict the response for test dataset
y_pred = knn.predict(X_test)

print(y_pred)
```

```
[1 1 0 0 0 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 0 1 0 1
 1 0 0 1 1 0 1 1 1 0 0 1 1 0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 0 1 0 0 1 1 1 0
 1 0 0 0 1 1 1 0 0 0 1 1 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 1
 1 0 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 0 1 1 1 1 0 1 1 0 0 0 1 0 1 0 1 1 0 1 0
 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 0 0 0 1 1 0 1 1 1 1 0 1 0 1 1
 1 0 1 1 1 0 0 1 1 1 0 0 1 1 0 1 1 0 1 0 1 0 0 1 1 1 1 0 1 1 1 0 1 1 1 1 1 0
 1 0 1 0 1 1]
```

In [32]:
```
print (metrics.accuracy_score(y_test, y_pred))
```

```
0.9166666666666666
```

In [27]:
```
#Finding a the best value for K
#For to to calculate K from 1 through 25 and record testing accuracy

k_range = range(1,25)
scores = []
for k in k_range:
    knn = KNeighborsClassifier(n_neighbors = k)
    knn.fit(X_train, y_train)
    y_pred = knn.predict(X_test)
    scores.append(metrics.accuracy_score(y_test,y_pred))
```

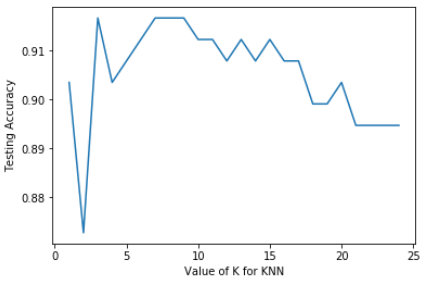
In [28]:
```
# Plot relationship between K and testing accuracy

import matplotlib.pyplot as plt

%matplotlib inline

plt.plot(k_range, scores)
plt.xlabel('Value of K for KNN')
plt.ylabel('Testing Accuracy')
```

Out[28]:
```
Text(0,0.5,'Testing Accuracy')
```



In [ ]: