Classification-Breast-Cancer

June 8, 2021

1 Classification - Breast Cancer

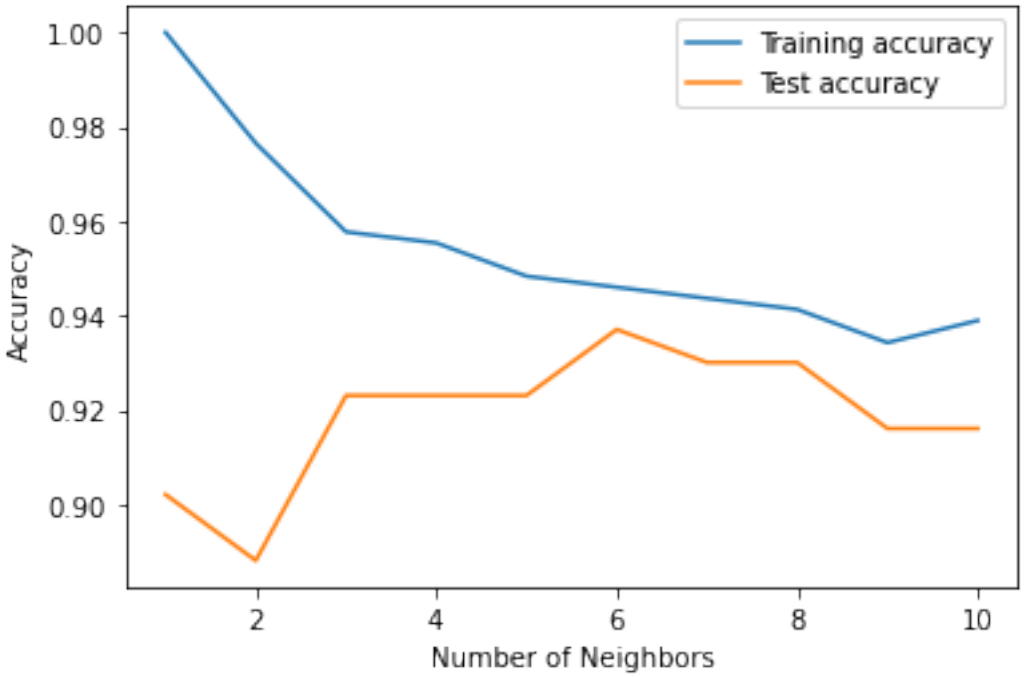
We demonstrate how increasing the number of neighbors in kNN classification affects the model's prediction accuracy.

[2]:
```
from sklearn.datasets import load_breast_cancer
cancer = load_breast_cancer()
```

[3]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
        cancer.data, cancer.target, stratify=cancer.target, random_state=66)
```

[6]:
```
from sklearn.neighbors import KNeighborsClassifier
training_accuracy = []
test_accuracy = []
#try n_neighbors from 1 to 10
neighbors_settings = range(1, 11)
for n_neighbors in neighbors_settings:
    # build the model
    clf = KNeighborsClassifier(n_neighbors=n_neighbors)
    clf.fit(X_train, y_train)
    #record training set accuracy
    training_accuracy.append(clf.score(X_train, y_train))
    # record generalisation accuracy
    test_accuracy.append(clf.score(X_test, y_test))
```

[10]:
```
import matplotlib.pyplot as plt
plt.plot(neighbors_settings, training_accuracy, label="Training accuracy")
plt.plot(neighbors_settings, test_accuracy, label="Test accuracy")
plt.ylabel("Accuracy")
plt.xlabel("Number of Neighbors")
plt.legend()
plt.show()
```


Training accuracy
Test accuracy
1.00
0.98
0.96
0.94
0.92
0.90
Accuracy
2
4
6
8
10
Number of Neighbors