KNN from Scratch

December 10, 2022

[1]:
```
from sklearn import datasets
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from collections import Counter
```

[2]:
```
data = datasets.load_breast_cancer()
```

[3]:
```
X_train, X_test, Y_train, Y_test = train_test_split(data.data, data.target)
```

0.1 Knn from scratch

[7]:
```
# lazy learner in train it do nothing
def train(X, Y):
    return

def predict_one(x_train, y_train, x_test, k):
    distances = []
    for i in range(len(x_train)):
        distance = ((x_train[i, :] - x_test)**2).sum()
        distances.append([distance, i])
    distances = sorted(distances)
    targets = []
    for i in range(k):
        index_of_training_data = distances[i][1]
        targets.append(y_train[index_of_training_data])
    return Counter(targets).most_common(1)[0][0]

def predict(X_train, Y_train, X_test_full, K):
    predictions = []
    for i in X_test_full:
        p = predict_one(X_train, Y_train, i, K)
        predictions.append(p)
    return predictions
```

1 knn scratch score

[5]:
```
Y_pred = predict(X_train, Y_train, X_test, 5)
accuracy_score(Y_pred, Y_test)
```

[5]:
```
0.9440559440559441
```

1.1 Inbuilt KNN score

[6]:
```
clf = KNeighborsClassifier()
clf.fit(X_train, Y_train)
Y_pred = clf.predict(X_test)
accuracy_score(Y_pred, Y_test)
```

[6]:
```
0.9440559440559441
```