Luke Scott

COSC 311

Homework 3

Dr. Wang

Lab Report

Source Code

Homework3.py

```
import matplotlib.pyplot as plt
import seaborn as sns; sns.set()  # for plot styling
import numpy as np
from sklearn.tree import DecisionTreeClassifier
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from warnings import simplefilter
from sklearn.neighbors import KNeighborsClassifier



simplefilter(action='ignore', category=FutureWarning)


data = pd.read_csv('FoodTypeDataset.csv',
                   names=['v1','v2','v3','v4','v5','v6','v7','v8','v9','v10',
                          'v11','v12','v13','v14','v15','v16','v17','v18','v19','v20',
                          'v21','v22','v23','v24','v25','v26','v27','v28','v29','v30',
                          'v31','v32','v33','v34','v35','v36','v37','v38','v39','v40',
                          'v41','v42','v43','v44','v45','v46','v47','v48','v49','v50',
                          'v51','v52','v53','v54','v55','v56','v57','v58','v59','v60',
```

```
                    'v61','v62','v63','v64','target'])

cols = ['v1','v2','v3','v4','v5','v6','v7','v8','v9','v10',
      'v11','v12','v13','v14','v15','v16','v17','v18','v19','v20',
      'v21','v22','v23','v24','v25','v26','v27','v28','v29','v30',
      'v31','v32','v33','v34','v35','v36','v37','v38','v39','v40',
      'v41','v42','v43','v44','v45','v46','v47','v48','v49','v50',
      'v51','v52','v53','v54','v55','v56','v57','v58','v59','v60',
      'v61','v62','v63','v64']

x = data[cols].values

y = data['target'].values

X_train, X_test, Y_train, Y_test = train_test_split(x, y, test_size=.2, random_state=0)

knn = KNeighborsClassifier(n_neighbors=7)
knn.fit(X_train, Y_train)
knn_pred = knn.predict(X_test)
print(knn.score(X_test, Y_test))

cm = confusion_matrix(Y_test, knn_pred)
print(cm)

mat = confusion_matrix(Y_test, knn_pred)
sns.heatmap(mat.T, square=True, annot=True, fmt='d', cbar=False,
          xticklabels=range(1, 21),
          yticklabels=range(1, 21))
```

```
plt.xlabel('true label')
plt.ylabel('predicted label')


print(classification_report(Y_test, knn_pred))
```

Notes

1. I chose to use the KNN model for this example, at first I tried to test different parameters with the Decision Tree Model, however I could not get a score over 35%. This is a very complex set of data so I was unable to get a great score in general, however the KNN model with 7 neighbors was able to score around a 55%
2. The confusion matrix for this model shows a high misclassification between 7 and 12. Most other classifications tend to be somewhat accurate with a few scattered outliers. It is important to note that no prediction was 100% accurate, the closest however was the model's prediction of food label 9, which classified 11 out of 13 samples correctly.