A)

from pandas import read\_excel

from random import sample

from sklearn.model\_selection import train\_test\_split

from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import classification\_report, confusion\_matrix

import matplotlib.pyplot as plt

def read\_data():

table = read\_excel("P4Data.xlsx")

data\_sets = []

classes = []

for x, y, c in zip(table['X'], table['Y'], table['Class']):

data\_sets.append([x, y])

classes.append([c])

return (data\_sets, classes)

def get\_testing\_indicies(training\_indices, data\_size):

testing\_indicies = []

for index in range(data\_size):

if index not in training\_indices:

testing\_indicies.append(index)

return testing\_indicies

def build\_50\_by\_50():

lst = []

for i in range(1, 51):

for j in range(1, 51):

lst.append([i, j])

return lst

def create\_graph(coords, pred):

pred0x = []

pred0y = []

pred1x = []

pred1y = []

for i in range(len(pred)):

if pred[i] == 0:

pred0x.append(coords[i][0])

pred0y.append(coords[i][1])

else:

pred1x.append(coords[i][0])

pred1y.append(coords[i][1])

plt.plot(pred0x, pred0y, 'ks', pred1x, pred1y, 'rs')

plt.xlabel('x')

plt.ylabel('y')

plt.title('Decision Tree Predictions')

plt.axis([0, 50, 0, 50])

plt.show()

def main():

(data\_sets, classes) = read\_data()

x\_train, x\_test, y\_train, y\_test = train\_test\_split(

data\_sets, classes, test\_size=.30)

classifier = DecisionTreeClassifier()

classifier.fit(x\_train, y\_train)

y\_pred = classifier.predict(x\_test)

print(confusion\_matrix(y\_test, y\_pred))

print(classification\_report(y\_test, y\_pred))

n\_test = build\_50\_by\_50()

n\_pred = classifier.predict(n\_test)

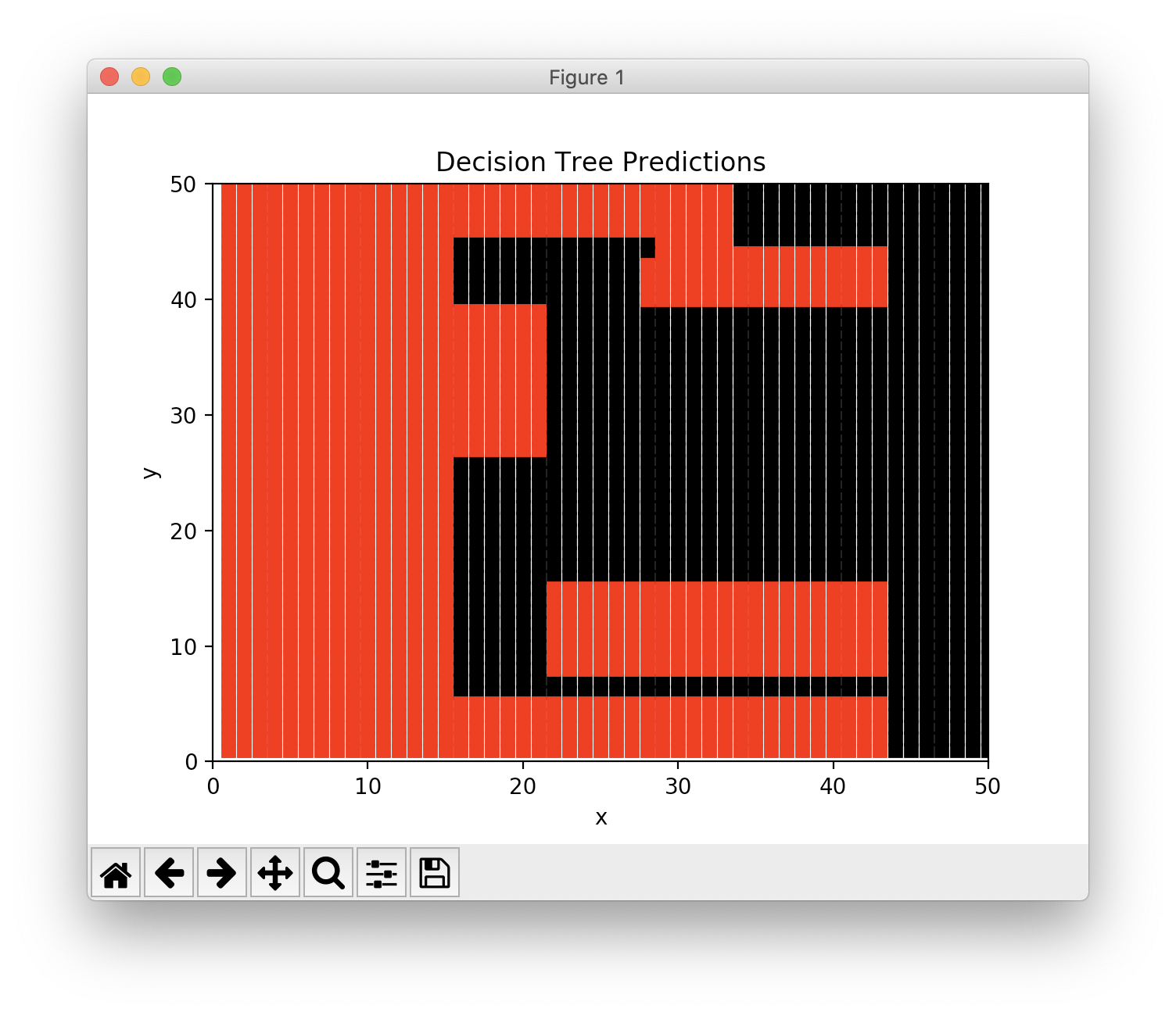
create\_graph(n\_test, n\_pred)

if \_\_name\_\_ == "\_\_main\_\_":

main()

A screenshot of a cell phone

Description automatically generated



In the graph above, the black points represent coordinate points that the classifier predicted to be in class 0 and the red points are those predicted to be in class 1. Generally, class 0 datapoints seem to appear closer to x = 50 and around the middle section of y values (~25 +/- 10). The class 1 data points seem to trend towards the x < 15 area and y < 5 area with some other small areas in the “L” shape that it makes. I think that there the class 0 boundary at y = 6 should not be there and it should just be an entirely red block for x = 21 through x = 42 and y = 0 through y = 15.

B)

A screenshot of a computer

Description automatically generated