Teddie Davis tdavis08@syr.edu Assignment 4

You pick 1 value than compare it to all other values, then go to the 2nd one and compare it to all the other values after that.

Question 1

a) The closest pairs for Euclidean distance are (4.9, 3.1, 1.5, 0.1) and (4.9, 3.1, 1.5, 0.1) with Euclidean distance of 0.0

```
import csv
import numpy as np

iris = []

with open('iris.data', newline='') as csvfile:
    spamreader = csv.reader(csvfile, delimiter=',')
    for row in spamreader:
        p = (float(row[0]), float(row[1]), float(row[2]), float(row[3]))
        iris.append(p)

lowestValue = 99.99
matchingPair = []

for x in range(len(iris)):
    point1 = np.array(iris[x])

    for i in range(x, len(iris)):
        if x != i:
            point2 = np.array(iris[i])
            dist = np.linalg.norm(point1 - point2)

            if lowestValue is None or lowestValue > dist :
                lowestValue = dist
                matchingPair = (x, i)
                print(matchingPair)

print("The closest pairs for Euclidean distance are {0} and {1} with Euclidean distance of {2}".format(iris[matchingPair[0]], iris[matchingPair[1]], lowestValue ))
```

b) The closest pairs for Mahattan distance are (4.9, 3.1, 1.5, 0.1) and (4.9, 3.1, 1.5, 0.1) with distance of 0.0

```
import csv
import numpy as np

iris = []

with open('iris.data', newline='') as csvfile:
    spamreader = csv.reader(csvfile, delimiter=',')
    for row in spamreader:
        p = (float(row[0]), float(row[1]), float(row[2]), float(row[3]))
        iris.append(p)

lowestValue = 99.99
matchingPair = []

for x in range(len(iris)):
    point1 = np.array(iris[x])
    #print(point1[0])
    for i in range(x, len(iris)):
        if x != i:
            point2 = np.array(iris[i])
            dist = abs(point1[0] - point2[0]) + abs(point1[1] - point2[1]) +
abs(point1[2] - point2[2]) + abs(point1[3] - point2[3])

            #dist = np.linalg.norm(point1 - point2)

            if lowestValue is None or lowestValue > dist :
                lowestValue = dist
                matchingPair = (x, i)
                print(matchingPair)

print("The closest pairs for Mahattan distance are {0} and {1} with distance of
{2}".format(iris[matchingPair[0]], iris[matchingPair[1]], lowestValue ))
```

c) The closest pairs for cosine similarity (4.9, 3.1, 1.5, 0.1) and (4.9, 3.1, 1.5, 0.1) with distance of 1.0

```
import csv
import numpy as np


iris = []

with open('iris.data', newline='') as csvfile:
    spamreader = csv.reader(csvfile, delimiter=',')
    for row in spamreader:
        p = (float(row[0]), float(row[1]), float(row[2]), float(row[3]))
        iris.append(p)

lowestValue = -99.99
matchingPair = []


for x in range(len(iris)):
    point1 = np.array(iris[x])

    for i in range(x, len(iris)):
        if x != i:
            point2 = np.array(iris[i])

            dist = np.dot(point1, point2)/(np.linalg.norm(point1)*np.linalg.norm(point2))


            if lowestValue is None or lowestValue < dist :
                lowestValue = dist
                matchingPair = (x, i)

print("The closest pairs for cosine similarity {0} and {1} with distance of
{2}".format(iris[matchingPair[0]], iris[matchingPair[1]], lowestValue ))
```

Question 2

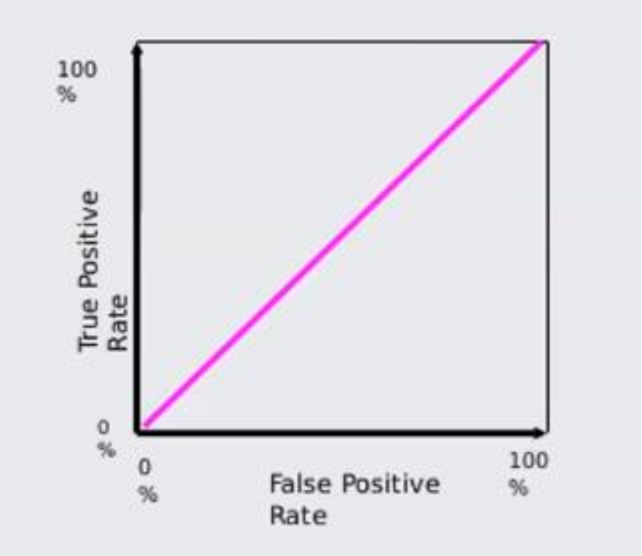
Seen Excel file to see work

| | |
|---|---|
| Input 1 | Not Red, No Leaves, Has seeds<br>0.092592593 |
| Input 2 | Red, No Leaves, No seeds<br>0.185185185 |
| Input 3 | Red, Has Leaves, Has Seeds<br>0.12962963 |

Question 3 a)	A point on the curve corresponds to an sensitivity/specificity pair to the decision threshold.

Question 3 b) “A” – has a better AUC percent than b. Also “A” has an better overall true positive and false Positive rate.

Question 3 c)  Random guess would look somewhat like an diagonal line. It wouldn’t be a perfect diagonal line and wouldn’t have a lot of variation and would not be very straight.



Question 4)

DCF = Cost(FR) * P(client) * FRR + Cost(FA) * P(imposter) FAR

P(client) = .5

P(Imposter) = .5

Cost(FR) $100

Cost(FA) $10

I would go with the “V-norm” algorithm. I would go with this one because if we falsely reject a customer, we lost lose that customer and “V-norm” has the lowest starting false reject rate. Based on the given information if we falsely accept someone, the user who we think it is gets an email with a way to reject the charges. Base on not having any information about profit lose I assumed the possible lose profit would be $10. This would account for someone getting to park then the customer being able to reject the charge.

Question 5)

Subtract every value by the smallest value and divide by .

|  |  |  |  | subtract | divide by range |
|---|---|---|---|---|---|
| L' | 8.6 | -8.6 | = | 0 | 0 |
| P | 8.9 | -8.6 | = | 0.3 | 0.15 |
| P' | 9.1 | -8.6 | = | 0.5 | 0.25 |
| Q | 9.3 | -8.6 | = | 0.7 | 0.35 |
| R | 9.5 | -8.6 | = | 0.9 | 0.45 |
| S | 9.6 | -8.6 | = | 1 | 0.5 |
| S' | 9.9 | -8.6 | = | 1.3 | 0.65 |
| T | 10.4 | -8.6 | = | 1.8 | 0.9 |
| T' | 10.6 | -8.6 | = | 2 | 1 |

After subtracting R from Q, R from P’, R from L’, R from S, R from S’, R from T and R from T’ the values are

|  |  |
|---|---|
| RQ | 0.415 |
| RP' | 0.2 |
| RP | 0.3 |
| RL’ | 0.45 |
| RS | 0.05 |
| RS’ | 0.2 |
| RT | 0.45 |
| RT' | 0.55 |