**Assignment 4**

By Harsh Sharma

P5. (Distance-based outlier detection) For this practice, you will use the iris data set. For this distance-based outlier detection, you will use k-nearest neighbor information.

(1) Compute the distance from each data object to its k-nearest neighbor (k=1)

(2) Report the data objects with top 5 largest distances from (1) as outliers

(3) Compute the distance from each data object to its k-nearest neighbor (k=3)

(4) Report the data objects with top 5 largest distances from (3) as outliers

(5) If any, list the data objects in both lists from (2) and (4) as outliers

Next, we will use the number of neighboring objects of each object in a given neighbor distance for outlier detection.

(6) Count the number of neighboring objects per each object within a distance threshold, 2

(7) Report the data objects with top 5 smallest neighbors as outliers

**Solution:**

The solution is the R file called ‘P5 Solution.R’. Here are the answers.

(1)

print(nearest\_neighbor\_distance)

1 2 3 4 5 6 7 8

0.1000000 0.1414214 0.1414214 0.1414214 0.1414214 0.3316625 0.2236068 0.1000000

9 10 11 12 13 14 15 16

0.1414214 0.1000000 0.1000000 0.2236068 0.1414214 0.2449490 0.4123106 0.3605551

17 18 19 20 21 22 23 24

0.3464102 0.1000000 0.3316625 0.1414214 0.2828427 0.1414214 0.4582576 0.2000000

25 26 27 28 29 30 31 32

0.3000000 0.1732051 0.2000000 0.1414214 0.1414214 0.1414214 0.1414214 0.2828427

33 34 35 36 37 38 39 40

0.3464102 0.3464102 0.1000000 0.2236068 0.3000000 0.1414214 0.1414214 0.1000000

41 42 43 44 45 46 47 48

0.1414214 0.6244998 0.2000000 0.2236068 0.3605551 0.1414214 0.1414214 0.1414214

49 50 51 52 53 54 55 56

0.1000000 0.1414214 0.2645751 0.2645751 0.2645751 0.2000000 0.2449490 0.3000000

57 58 59 60 61 62 63 64

0.2645751 0.1414214 0.2449490 0.3872983 0.3605551 0.3000000 0.4898979 0.1414214

65 66 67 68 69 70 71 72

0.4242641 0.1414214 0.2000000 0.2449490 0.2645751 0.1732051 0.2236068 0.3316625

73 74 75 76 77 78 79 80

0.3605551 0.2236068 0.2000000 0.1414214 0.3162278 0.3162278 0.2000000 0.3464102

81 82 83 84 85 86 87 88

0.1414214 0.1414214 0.1414214 0.3316625 0.2000000 0.3741657 0.2828427 0.2645751

89 90 91 92 93 94 95 96

0.1732051 0.2000000 0.2645751 0.1414214 0.1414214 0.1414214 0.1732051 0.1414214

97 98 99 100 101 102 103 104

0.1414214 0.2000000 0.3872983 0.1414214 0.4242641 0.0000000 0.3872983 0.2449490

105 106 107 108 109 110 111 112

0.3000000 0.2645751 0.7348469 0.2645751 0.5567764 0.6324555 0.2236068 0.3464102

113 114 115 116 117 118 119 120

0.1732051 0.2645751 0.4898979 0.3000000 0.1414214 0.4123106 0.4123106 0.4358899

121 122 123 124 125 126 127 128

0.2236068 0.3162278 0.2645751 0.1732051 0.3000000 0.3464102 0.1732051 0.1414214

129 130 131 132 133 134 135 136

0.1000000 0.3464102 0.2645751 0.4123106 0.1000000 0.3316625 0.5385165 0.5385165

137 138 139 140 141 142 143 144

0.2449490 0.1414214 0.1414214 0.1732051 0.2449490 0.2449490 0.0000000 0.2236068

145 146 147 148 149 150

0.2449490 0.2449490 0.2449490 0.2236068 0.2449490 0.2828427

(2)

> print("Top 5 Outliers:")

[1] "Top 5 Outliers:"

> print(top\_5\_outliers)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

107 4.9 2.5 4.5 1.7 virginica

110 7.2 3.6 6.1 2.5 virginica

42 4.5 2.3 1.3 0.3 setosa

109 6.7 2.5 5.8 1.8 virginica

135 6.1 2.6 5.6 1.4 virginica

(3)

print(k\_nearest\_distances)

1 2 3 4 5 6 7 8

[1,] 0.1000000 0.1414214 0.1414214 0.1414214 0.1414214 0.3316625 0.2236068 0.1000000

[2,] 0.1414214 0.1414214 0.2449490 0.1732051 0.1414214 0.3464102 0.2645751 0.1414214

[3,] 0.1414214 0.1414214 0.2645751 0.2236068 0.1732051 0.3605551 0.3000000 0.1732051

9 10 11 12 13 14 15 16

[1,] 0.1414214 0.1000000 0.1000000 0.2236068 0.1414214 0.2449490 0.4123106 0.3605551

[2,] 0.3000000 0.1732051 0.2828427 0.2236068 0.1732051 0.3162278 0.4690416 0.5477226

[3,] 0.3162278 0.1732051 0.3000000 0.2828427 0.2000000 0.3464102 0.5477226 0.6164414

17 18 19 20 21 22 23 24

[1,] 0.3464102 0.1000000 0.3316625 0.1414214 0.2828427 0.1414214 0.4582576 0.2000000

[2,] 0.3605551 0.1414214 0.3872983 0.1414214 0.3000000 0.2449490 0.5099020 0.2645751

[3,] 0.3872983 0.1732051 0.4690416 0.2449490 0.3605551 0.2449490 0.5099020 0.3741657

25 26 27 28 29 30 31 32

[1,] 0.3000000 0.1732051 0.2000000 0.1414214 0.1414214 0.1414214 0.1414214 0.2828427

[2,] 0.3741657 0.2000000 0.2236068 0.1414214 0.1414214 0.1732051 0.1414214 0.3000000

[3,] 0.4123106 0.2236068 0.2236068 0.1414214 0.1414214 0.2236068 0.1732051 0.3000000

33 34 35 36 37 38 39 40

[1,] 0.3464102 0.3464102 0.1000000 0.2236068 0.3000000 0.1414214 0.1414214 0.1000000

[2,] 0.3464102 0.3605551 0.1414214 0.3000000 0.3162278 0.2449490 0.2000000 0.1414214

[3,] 0.3741657 0.3872983 0.1414214 0.3162278 0.3316625 0.2645751 0.2449490 0.1414214

41 42 43 44 45 46 47 48

[1,] 0.1414214 0.6244998 0.2000000 0.2236068 0.3605551 0.1414214 0.1414214 0.1414214

[2,] 0.1732051 0.7141428 0.2236068 0.2645751 0.3741657 0.2000000 0.2449490 0.1414214

[3,] 0.1732051 0.7681146 0.3000000 0.3162278 0.4123106 0.2000000 0.2449490 0.2236068

49 50 51 52 53 54 55 56

[1,] 0.1000000 0.1414214 0.2645751 0.2645751 0.2645751 0.2000000 0.2449490 0.3000000

[2,] 0.2236068 0.1732051 0.3316625 0.3162278 0.2828427 0.3000000 0.3162278 0.3162278

[3,] 0.2449490 0.2236068 0.4358899 0.3464102 0.3162278 0.3162278 0.3741657 0.3162278

57 58 59 60 61 62 63 64

[1,] 0.2645751 0.1414214 0.2449490 0.3872983 0.3605551 0.3000000 0.4898979 0.1414214

[2,] 0.3741657 0.3872983 0.2449490 0.5099020 0.4582576 0.3316625 0.5196152 0.2236068

[3,] 0.4242641 0.4582576 0.3162278 0.5196152 0.6708204 0.3605551 0.5477226 0.2449490

65 66 67 68 69 70 71 72

[1,] 0.4242641 0.1414214 0.2000000 0.2449490 0.2645751 0.1732051 0.2236068 0.3316625

[2,] 0.4472136 0.3162278 0.3000000 0.2828427 0.5099020 0.2449490 0.3000000 0.3464102

[3,] 0.5099020 0.3162278 0.3872983 0.3316625 0.5385165 0.2645751 0.3605551 0.3741657

73 74 75 76 77 78 79 80

[1,] 0.3605551 0.2236068 0.2000000 0.1414214 0.3162278 0.3162278 0.2000000 0.3464102

[2,] 0.3605551 0.3000000 0.2645751 0.2449490 0.3464102 0.3741657 0.2449490 0.4242641

[3,] 0.4123106 0.3872983 0.3605551 0.2645751 0.3464102 0.4123106 0.3316625 0.4358899

81 82 83 84 85 86 87 88

[1,] 0.1414214 0.1414214 0.1414214 0.3316625 0.2000000 0.3741657 0.2828427 0.2645751

[2,] 0.1732051 0.2645751 0.2645751 0.3605551 0.4123106 0.4242641 0.3162278 0.5744563

[3,] 0.3000000 0.3464102 0.2828427 0.3605551 0.4795832 0.4582576 0.3162278 0.5916080

89 90 91 92 93 94 95 96

[1,] 0.1732051 0.200000 0.2645751 0.1414214 0.1414214 0.1414214 0.1732051 0.1414214

[2,] 0.1732051 0.244949 0.3162278 0.2000000 0.2449490 0.3605551 0.2236068 0.1732051

[3,] 0.2236068 0.300000 0.4242641 0.3000000 0.2645751 0.3872983 0.2645751 0.2449490

97 98 99 100 101 102 103 104

[1,] 0.1414214 0.2000000 0.3872983 0.1414214 0.4242641 0.0000000 0.3872983 0.2449490

[2,] 0.1414214 0.3316625 0.3872983 0.1732051 0.5000000 0.2645751 0.4000000 0.2449490

[3,] 0.1732051 0.3464102 0.7211103 0.2236068 0.5099020 0.3162278 0.4123106 0.3316625

105 106 107 108 109 110 111 112

[1,] 0.3000000 0.2645751 0.7348469 0.2645751 0.5567764 0.6324555 0.2236068 0.3464102

[2,] 0.3162278 0.5291503 0.7615773 0.4358899 0.6000000 0.6708204 0.3741657 0.3741657

[3,] 0.3605551 0.5477226 0.7937254 0.5291503 0.6164414 0.7071068 0.4242641 0.3741657

113 114 115 116 117 118 119 120

[1,] 0.1732051 0.2645751 0.4898979 0.3000000 0.1414214 0.4123106 0.4123106 0.4358899

[2,] 0.3464102 0.2645751 0.5099020 0.3741657 0.2449490 0.8185353 0.5477226 0.5196152

[3,] 0.3605551 0.3316625 0.5099020 0.3741657 0.3605551 0.8602325 0.8944272 0.5385165

121 122 123 124 125 126 127 128

[1,] 0.2236068 0.3162278 0.2645751 0.1732051 0.3000000 0.3464102 0.1732051 0.1414214

[2,] 0.2645751 0.3162278 0.4123106 0.2449490 0.3162278 0.3872983 0.2449490 0.2449490

[3,] 0.3000000 0.3316625 0.6082763 0.3605551 0.3741657 0.4358899 0.2828427 0.2828427

129 130 131 132 133 134 135 136

[1,] 0.1000000 0.3464102 0.2645751 0.4123106 0.1000000 0.3316625 0.5385165 0.5385165

[2,] 0.3162278 0.5099020 0.4582576 0.8831761 0.3000000 0.3605551 0.5567764 0.5477226

[3,] 0.3316625 0.5196152 0.4690416 0.9273618 0.4242641 0.3741657 0.5830952 0.6633250

137 138 139 140 141 142 143 144

[1,] 0.2449490 0.1414214 0.1414214 0.1732051 0.2449490 0.2449490 0.0000000 0.2236068

[2,] 0.3872983 0.2449490 0.2236068 0.3605551 0.2645751 0.3605551 0.2645751 0.3162278

[3,] 0.4242641 0.3872983 0.2828427 0.3605551 0.3464102 0.4690416 0.3162278 0.3162278

145 146 147 148 149 150

[1,] 0.2449490 0.2449490 0.2449490 0.2236068 0.2449490 0.2828427

[2,] 0.3000000 0.3605551 0.3741657 0.3464102 0.3000000 0.3162278

[3,] 0.3162278 0.3605551 0.3872983 0.3605551 0.5567764 0.3316625

(4)

> print("Top 5 Outliers:")

[1] "Top 5 Outliers:"

> print(top\_5\_outliers)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

3 4.7 3.2 1.3 0.2 setosa

2 4.9 3.0 1.4 0.2 setosa

1 5.1 3.5 1.4 0.2 setosa

NA NA NA NA NA <NA>

NA.1 NA NA NA NA <NA>

(5)

[1] "No data objects found in both lists as outliers."

(6)

print(neighboring\_objects\_count)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

49 50 49 50 49 49 49 50 49 50 49 50 50 48 48 47 49 49 49 49 50 49

23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44

49 52 52 51 50 49 49 50 50 50 49 48 50 49 49 49 49 50 49 47 49 50

45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66

50 50 49 49 49 49 87 87 89 70 88 83 89 38 88 65 38 85 70 90 66 87

67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88

86 73 88 67 86 79 88 90 89 87 90 91 90 62 65 60 72 88 83 86 90 85

89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110

76 71 78 90 76 37 78 79 80 86 54 78 68 88 69 87 75 37 72 55 78 49

111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132

88 90 81 87 86 85 89 28 25 89 73 87 32 90 75 68 89 89 85 71 60 30

133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150

84 90 88 43 76 88 89 80 75 82 88 69 72 84 90 89 80 89

>

(7)

print("Top 5 Outliers:")

[1] "Top 5 Outliers:"

> print(top\_5\_outliers)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

119 7.7 2.6 6.9 2.3 virginica

118 7.7 3.8 6.7 2.2 virginica

132 7.9 3.8 6.4 2.0 virginica

123 7.7 2.8 6.7 2.0 virginica

94 5.0 2.3 3.3 1.0 versicolor