**Assignment 4**

By Harsh Sharma

P4. (Density-based local outlier detection) For this practice, you will use the iris data set. LOF (Local Outlier Factor) is an algorithm for identifying density-based local outliers.

(1) Calculate the local outlier factor value of each object as its outlier score.

(2) Report the data objects of top 5 LOF values as outliers

(3) Compute kernel density estimates with the LOF values and generate the plot diagram for plotting of them.

**Solution:**

The solution is the R file called ‘P4 Solution.R’. Here are the answers.

(1)

print(iris$LOF)

[1] 0.9624149 0.9980281 1.1244614 0.9473478 1.0768750 1.1900530 0.9909691 0.9954913

[9] 1.1702821 0.9607724 1.2542188 1.1729924 0.9980281 1.1523279 1.1281174 1.2485819

[17] 1.1829703 1.0027215 1.3346907 1.0656077 1.5902613 1.1922665 2.1077309 1.5108668

[25] 1.4669203 1.0308436 1.3188356 0.9291017 0.9291017 0.9972336 0.9728637 1.5292459

[33] 1.2686264 1.1148585 0.9890112 1.3472101 1.3459493 1.2727643 1.2009936 1.0371338

[41] 1.0380848 2.4064854 1.0140464 1.2674228 1.3281089 1.0208344 1.0656077 1.0739280

[49] 1.1938117 1.1978260 1.0931981 1.0104093 1.0607719 0.9795255 0.9694234 1.1018860

[57] 1.1003782 1.2378798 1.0192553 1.4804385 1.4149410 1.1160331 1.7176875 1.0186381

[65] 1.4925739 0.9214874 1.1129799 1.0380017 1.3060016 0.9653809 0.9803214 1.1324987

[73] 1.0399103 1.0520400 0.9886459 1.0237249 0.9489064 1.0449801 0.9954574 1.2050696

[81] 1.0561252 1.0621708 0.9692484 1.0036211 1.3565734 1.1691020 0.9983495 1.3586172

[89] 0.9562342 0.9893589 1.1542840 1.0186381 0.9692484 1.2378798 0.9353166 0.9562342

[97] 1.0387712 1.0032700 1.1601034 1.0157126 1.2839935 1.0706585 1.0569548 1.0057564

[105] 1.0373544 1.1214264 1.9922988 0.9781150 1.4995758 1.8402443 0.9734784 0.9747798

[113] 0.9517462 0.9521860 1.2364854 1.0726422 0.9988396 1.2598646 1.2261373 1.2896320

[121] 1.0425966 1.0706585 1.1429680 1.0531339 0.9681929 1.0462010 0.9204444 1.0529793

[129] 1.0146323 1.0353673 1.0206477 1.3148779 1.0334960 1.0521490 1.4965556 1.1445097

[137] 1.1345143 1.0102623 1.0384913 1.0469830 0.9911082 1.0985050 1.0706585 1.0017035

[145] 0.9772474 1.0175722 1.0190350 0.9954005 1.1568777 0.8306547

>

(2)

print(top\_5\_outliers)

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

42 4.5 2.3 1.3 0.3 setosa 2.406485

23 4.6 3.6 1.0 0.2 setosa 2.107731

107 4.9 2.5 4.5 1.7 virginica 1.992299

110 7.2 3.6 6.1 2.5 virginica 1.840244

63 6.0 2.2 4.0 1.0 versicolor 1.717688

>

(3)

print(lof\_density)

Call:

density.default(x = iris$LOF)

Data: iris$LOF (150 obs.); Bandwidth 'bw' = 0.0489

x y

Min. :0.6839 Min. :0.000618

1st Qu.:1.1513 1st Qu.:0.037989

Median :1.6186 Median :0.057335

Mean :1.6186 Mean :0.534440

3rd Qu.:2.0859 3rd Qu.:0.585489

Max. :2.5532 Max. :3.626024

A graph of a number of individuals

Description automatically generated with medium confidence