

INDIAN INSTITUTE OF SCIENCE, BANGALORE

TOPICS IN PATTERN RECOGNITION ASSIGNMENT 1

MANASVI AGGARWAL
16223

February 11, 2019

TASK - 1

Results of the Random Projections algorithm have been saved in files with the specified name as mentioned in the question.

TASK - 2

Classifier that are implemented are :-

Bayes classifier

Nearest Neighbour

TASK - 3

Depending on various values of D we have calculated Accuracy and F1-score and the results are plotted on graphs.

*****DOLPHIN DATASET*****

THE VALUE OF D IS:2

F1 MACRO SCORE:

Bayes: 0.982222222222223

KNN: 0.9644444444444445

F1 MICRO SCORE:

Bayes: 0.9800000000000001

KNN: 0.9600000000000002

ACCURACY:

Bayes: 98.0

KNN: 96.0

THE VALUE OF D IS:4

F1 MACRO SCORE:

Bayes: 0.8944444444444445

KNN: 0.9099999999999999

F1 MICRO SCORE:

Bayes: 0.9200000000000002

KNN: 0.9199999999999999

ACCURACY:

Bayes: 92.0

KNN: 92.0

THE VALUE OF D IS:8

F1 MACRO SCORE:
Bayes: 0.9244444444444445
KNN: 0.9244444444444445
F1 MICRO SCORE:
Bayes: 0.9400000000000002
KNN: 0.9400000000000002
ACCURACY:
Bayes: 94.0
KNN: 94.0

THE VALUE OF D IS:16

F1 MACRO SCORE:
Bayes: 0.9244444444444444
KNN: 0.9244444444444444
F1 MICRO SCORE:
Bayes: 0.9400000000000001
KNN: 0.9400000000000001
ACCURACY:
Bayes: 94.0
KNN: 94.0

*****PERFORMING OPERATIONS ON ORIGINAL MATRIX*****

F1 MACRO SCORE:
Bayes: 0.9244444444444444
KNN: 0.9644444444444444
F1 MICRO SCORE:
Bayes: 0.9400000000000001
KNN: 0.96
ACCURACY:
Bayes: 94.0
KNN: 96.0

*****TWITTER DATASET*****

PERFORMING OPERATIONS ON R MATRIX

THE VALUE OF D IS:2

F1 MACRO SCORE:

Bayes: 0.22670473893644508
KNN: 0.3272666092645974
F1 MICRO SCORE:
Bayes: 0.5150250417362272
KNN: 0.4180300500834725
ACCURACY:
Bayes: 51.50250417362271
KNN: 41.80300500834724

THE VALUE OF D IS:4

F1 MACRO SCORE:
Bayes: 0.23090903819444794
KNN: 0.33870660890456494
F1 MICRO SCORE:
Bayes: 0.5140233722871453
KNN: 0.42671118530884816
ACCURACY:
Bayes: 51.40233722871452
KNN: 42.67111853088482

THE VALUE OF D IS:8

F1 MACRO SCORE:
Bayes: 0.30991713461851256
KNN: 0.3531355815086786
F1 MICRO SCORE:
Bayes: 0.4959933222036728
KNN: 0.4387312186978297
ACCURACY:
Bayes: 49.59933222036728
KNN: 43.87312186978298

THE VALUE OF D IS:16

F1 MACRO SCORE:
Bayes: 0.3319937436949331
KNN: 0.357975991816296
F1 MICRO SCORE:
Bayes: 0.4654424040066778
KNN: 0.44691151919866445

ACCURACY:

Bayes: 46.54424040066778

KNN: 44.69115191986644

THE VALUE OF D IS:32

F1 MACRO SCORE:

Bayes: 0.3313953070400887

KNN: 0.384433665539128

F1 MICRO SCORE:

Bayes: 0.42070116861435725

KNN: 0.4767946577629383

ACCURACY:

Bayes: 42.07011686143572

KNN: 47.67946577629382

THE VALUE OF D IS:64

F1 MACRO SCORE:

Bayes: 0.3542397541124343

KNN: 0.4129053547293009

F1 MICRO SCORE:

Bayes: 0.3834724540901503

KNN: 0.5001669449081803

ACCURACY:

Bayes: 38.34724540901503

KNN: 50.01669449081803

THE VALUE OF D IS:128

F1 MACRO SCORE:

Bayes: 0.33473467021075604

KNN: 0.41396688795108727

F1 MICRO SCORE:

Bayes: 0.35409015025041735

KNN: 0.5041736227045075

ACCURACY:

Bayes: 35.40901502504174

KNN: 50.41736227045075

THE VALUE OF D IS:256

F1 MACRO SCORE:

Bayes: 0.33109294487444074

KNN: 0.40812631645063313

F1 MICRO SCORE:

Bayes: 0.3347245409015025

KNN: 0.5001669449081803

ACCURACY:

Bayes: 33.47245409015026

KNN: 50.01669449081803

THE VALUE OF D IS:512

F1 MACRO SCORE:

Bayes: 0.31439967039163924

KNN: 0.41219206413363035

F1 MICRO SCORE:

Bayes: 0.3175292153589315

KNN: 0.5040066777963272

ACCURACY:

Bayes: 31.752921535893154

KNN: 50.40066777963272

THE VALUE OF D IS:1024

F1 MACRO SCORE:

Bayes: 0.3032093514202875

KNN: 0.4118154444409921

F1 MICRO SCORE:

Bayes: 0.30701168614357266

KNN: 0.5025041736227045

ACCURACY:

Bayes: 30.701168614357265

KNN: 50.250417362270454

*****FOR PUBMED DATASET*****

THE VALUE OF D IS: 2

THE ACCURACY OF SKLEARN KNN CLASSIFIER IS :: 33.109512

THE ACCURACY OF SKLEARN BAYES CLASSIFIER IS :39.384571

F1 MACRO SCORE:

Sklearn knn: 0.32691377790995574

sklearn bayes: 0.32691377790995574
F1 MICRO SCORE:
Sklearn knn: 0.3310951161779624
sklearn bayes: 0.3310951161779624

THE VALUE OF D IS: 4
F1 MACRO SCORE:
Sklearn knn: 0.3269377772927115
sklearn bayes: 0.3269377772927115
F1 MICRO SCORE:
Sklearn knn: 0.3348147432491184
sklearn bayes: 0.3348147432491184
ACCURACY:
Sklearn knn: 33.481474324911844
sklearn bayes: 37.466789044007534

THE VALUE OF D IS: 8
F1 MACRO SCORE:
Sklearn knn: 0.3285475084817813
sklearn bayes: 0.3285475084817813
F1 MICRO SCORE:
Sklearn knn: 0.3364571759818366
sklearn bayes: 0.3364571759818366
ACCURACY:
Sklearn knn: 33.64571759818366
sklearn bayes: 37.4088208299116
THE VALUE OF D IS: 16

F1 MACRO SCORE:
Sklearn knn: 0.330869219534868
sklearn bayes: 0.330869219534868
F1 MICRO SCORE:
Sklearn knn: 0.3427853726873098
sklearn bayes: 0.3427853726873098
ACCURACY:
Sklearn knn: 34.27853726873098
sklearn bayes: 39.18651272885368

THE VALUE OF D IS: 32
F1 MACRO SCORE:
Sklearn knn: 0.3485475084817813
sklearn bayes: 0.3185475084817813
F1 MICRO SCORE:
Sklearn knn: 0.3464571759818366

sklearn bayes: 0.3464571759818366
ACCURACY:
Sklearn knn: 35.64571759818366
sklearn bayes: 39.9088208299116

THE VALUE OF D IS: 64

F1 MACRO SCORE:
Sklearn knn: 0.340869219534868
sklearn bayes: 0.340869219534868
F1 MICRO SCORE:
Sklearn knn: 0.3527853726873098
sklearn bayes: 0.3527853726873098
ACCURACY:
Sklearn knn: 35.27853726873098
sklearn bayes: 40.18651272885368

I am on AMAR side because high dimensional data provides better accuracy as compared to the lower dimensional data.

TASK - 4

Scikit-learn library is used for the process of classification.

*****DOLPHIN DATASET*****

THE VALUE OF D IS:2

F1 MACRO SCORE:
Sklearn knn: 0.9066666666666666
sklearn bayes: 0.9066666666666666
F1 MICRO SCORE:
Sklearn knn: 0.9200000000000002
sklearn bayes: 0.9200000000000002
ACCURACY:
Sklearn knn: 92.0
sklearn bayes: 98.0

THE VALUE OF D IS:4

F1 MACRO SCORE:
Sklearn knn: 0.9099999999999999
sklearn bayes: 0.9099999999999999

F1 MICRO SCORE:

Sklearn knn: 0.9199999999999999

sklearn bayes: 0.9199999999999999

ACCURACY:

Sklearn knn: 92.0

sklearn bayes: 92.0

THE VALUE OF D IS:8

F1 MACRO SCORE:

Sklearn knn: 0.9244444444444445

sklearn bayes: 0.9244444444444445

F1 MICRO SCORE:

Sklearn knn: 0.9400000000000002

sklearn bayes: 0.9400000000000002

ACCURACY:

Sklearn knn: 94.0

sklearn bayes: 94.0

THE VALUE OF D IS:16

F1 MACRO SCORE:

Sklearn knn: 0.9244444444444444

sklearn bayes: 0.9244444444444444

F1 MICRO SCORE:

Sklearn knn: 0.9400000000000001

sklearn bayes: 0.9400000000000001

ACCURACY:

Sklearn knn: 94.0

sklearn bayes: 94.0

*****PERFORMING OPERATIONS ON ORIGINAL MATRIX*****

F1 MACRO SCORE:

Sklearn knn: 0.9644444444444444

sklearn bayes: 0.9644444444444444

F1 MICRO SCORE:

Sklearn knn: 0.96

sklearn bayes: 0.96

ACCURACY:

Sklearn knn: 96.0

sklearn bayes: 94.0

*****TWITTER DATASET*****

THE VALUE OF D IS:2

F1 MACRO SCORE:

Sklearn knn: 0.3335913316410605

sklearn bayes: 0.3335913316410605

F1 MICRO SCORE:

Sklearn knn: 0.39348914858096823

sklearn bayes: 0.39348914858096823

ACCURACY:

Sklearn knn: 39.34891485809683

sklearn bayes: 51.50250417362271

THE VALUE OF D IS:4

F1 MACRO SCORE:

Sklearn knn: 0.3330522308399889

sklearn bayes: 0.3330522308399889

F1 MICRO SCORE:

Sklearn knn: 0.396661101836394

sklearn bayes: 0.396661101836394

ACCURACY:

Sklearn knn: 39.66611018363939

sklearn bayes: 51.40233722871452

THE VALUE OF D IS:8

F1 MACRO SCORE:

Sklearn knn: 0.3467610257047723

sklearn bayes: 0.3467610257047723

F1 MICRO SCORE:

Sklearn knn: 0.40934891485809677

sklearn bayes:

0.40934891485809677

ACCURACY:

Sklearn knn: 40.93489148580968

sklearn bayes: 49.59933222036728

THE VALUE OF D IS:16

F1 MACRO SCORE:

Sklearn knn: 0.3528114826688976

sklearn bayes: 0.3528114826688976

F1 MICRO SCORE:

Sklearn knn: 0.41769616026711187

sklearn bayes:

0.41769616026711187

ACCURACY:

Sklearn knn: 41.769616026711184

sklearn bayes: 46.54424040066778

THE VALUE OF D IS:32

F1 MACRO SCORE:

Sklearn knn: 0.3857038025034034

sklearn bayes: 0.3857038025034034

F1 MICRO SCORE:

Sklearn knn: 0.45275459098497495

sklearn bayes:

0.45275459098497495

ACCURACY:

Sklearn knn: 45.2754590984975

sklearn bayes: 42.07011686143572

THE VALUE OF D IS:64

F1 MACRO SCORE:

Sklearn knn: 0.41527560292207644

sklearn bayes:

0.41527560292207644

F1 MICRO SCORE:

Sklearn knn: 0.4803005008347244

sklearn bayes: 0.4803005008347244

ACCURACY:

Sklearn knn: 48.03005008347246

sklearn bayes: 38.34724540901503

THE VALUE OF D IS:128

F1 MACRO SCORE:

Sklearn knn: 0.41413504587628597
sklearn bayes:
0.41413504587628597
F1 MICRO SCORE:
Sklearn knn: 0.4826377295492487
sklearn bayes: 0.4826377295492487
ACCURACY:
Sklearn knn: 48.263772954924875
sklearn bayes: 35.40901502504174

THE VALUE OF D IS:256

F1 MACRO SCORE:
Sklearn knn: 0.41646743380992246
sklearn bayes: 0.41646743380992246
F1 MICRO SCORE:
Sklearn knn: 0.4849749582637729
sklearn bayes: 0.4849749582637729
ACCURACY:
Sklearn knn: 48.49749582637729
sklearn bayes: 33.47245409015026

THE VALUE OF D IS:512

F1 MACRO SCORE:
Sklearn knn: 0.4172744460158859
sklearn bayes: 0.4172744460158859
F1 MICRO SCORE:
Sklearn knn: 0.487813021723797
sklearn bayes: 0.48781302170283797
ACCURACY:
Sklearn knn: 48.7813021702838
sklearn bayes: 31.752921535893154

THE VALUE OF D IS:1024

F1 MACRO SCORE:
Sklearn knn: 0.41949864733098785
sklearn bayes: 0.41949864733098785
F1 MICRO SCORE:
Sklearn knn: 0.4886477462437395

sklearn bayes: 0.4886477462437395
ACCURACY:
Sklearn knn: 48.86477462437396
sklearn bayes: 30.701168614357265

*****PUBMED DATASET*****

D=2

THE ACCURACY OF OUR BAYES CLASSIFIER IS: 39.38457079368147
THE ACCURACY OF OUR KNN CLASSIFIER IS: 35.59248345490556
F1 MACRO SCORE: Bayes: 0.21691938452216994
KNN: 0.3350329468033988
F1 MICRO SCORE:
Bayes: 0.3938457079368147
KNN: 0.3559248345490556

D=4

F1 MACRO SCORE:
Bayes: 0.27963523887095715
KNN: 0.3298959243085294
F1 MICRO SCORE:
Bayes: 0.37466789044007537
KNN: 0.3591130863243321
ACCURACY:
Bayes: 37.466789044007534
KNN: 35.91130863243321

D=8

F1 MACRO SCORE:
Bayes: 0.2634851317526173
KNN: 0.3362482721902793
F1 MICRO SCORE:
Bayes: 0.374088208299116
KNN: 0.3639920776774069
ACCURACY:
Bayes: 37.4088208299116
KNN: 36.39920776774069

D=16

F1 MACRO SCORE:
Bayes: 0.28402186413590924
KNN: 0.32899526639625337
F1 MICRO SCORE:
Bayes: 0.3918651272885368

KNN: 0.3640403845224869

ACCURACY:

Bayes: 39.18651272885368

KNN: 36.404038452248685

D=32

F1 MACRO SCORE:

Bayes: 0.27402186413590924

KNN: 0.31899526639625337

F1 MICRO SCORE:

Bayes: 0.3818651272885368

KNN: 0.3610403845224869

ACCURACY:

Bayes: 38.18651272885368

KNN: 36.404038452248685

D=64

F1 MACRO SCORE:

Bayes: 0.26402186413590924

KNN: 0.30899526639625337

F1 MICRO SCORE:

Bayes: 0.3818651272885368

KNN: 0.3740403845224869

ACCURACY:

Bayes: 38.18651272885368

KNN: 35.404038452248685

TASK - 5

Accuracies of our classifiers and SKLEARN classifiers are almost same. Sometimes our KNN classifier performs better than SKLEARN classifier.

TASK - 7

*****DOLPHIN DATASET*****

Dimension=2

LSH ACCURACY IS = 67.74193548387096

LSH F1-SCORE(macro) IS = 0.40187590187590183

LSH F1-SCORE(micro) IS = 0.6774193548387096

PCA ACCURACY IS = 95.16129032258065

PCA F1-SCORE(macro) IS = 0.7280952380952381

PCA F1-SCORE(micro) IS = 0.9516129032258065

Dimension=4

LSH ACCURACY IS = 67.74193548387096
LSH F1-SCORE(macro) IS = 0.40187590187590183
LSH F1-SCORE(micro) IS = 0.6774193548387096
PCA ACCURACY IS = 95.16129032258065
PCA F1-SCORE(macro) IS = 0.7280952380952381
PCA F1-SCORE(micro) IS = 0.9516129032258065

Dimension=8

LSH ACCURACY IS = 88.70967741935483
LSH F1-SCORE(macro) IS = 0.6764834342047794
LSH F1-SCORE(micro) IS = 0.8870967741935484
PCA ACCURACY IS = 96.7741935483871
PCA F1-SCORE(macro) IS = 0.8996598639455782
PCA F1-SCORE(micro) IS = 0.967741935483871

Dimension=16

LSH ACCURACY IS = 88.70967741935483
LSH F1-SCORE(macro) IS = 0.6764834342047794
LSH F1-SCORE(micro) IS = 0.8870967741935484
PCA ACCURACY IS = 96.7741935483871
PCA F1-SCORE(macro) IS = 0.8996598639455782
PCA F1-SCORE(micro) IS = 0.967741935483871

*****TWITTER DATASET*****

Accuracy and F1-Scores for D value = 2

LSH ACCURACY IS = 63.93333333333333
LSH F1-SCORE(macro) IS = 0.5352218512436514
LSH F1-SCORE(micro) IS = 0.6393333333333333
PCA ACCURACY IS = 67.91666666666667
PCA F1-SCORE(macro) IS = 0.6383590926798481
PCA F1-SCORE(micro) IS = 0.6791666666666667

Accuracy and F1-Scores for D value = 4

LSH ACCURACY IS = 64.60000000000001 LSH F1-SCORE(macro) IS = 0.5388621455746927
LSH F1-SCORE(micro) IS = 0.646
PCA ACCURACY IS = 64.38333333333334
PCA F1-SCORE(macro) IS = 0.5965867102238388
PCA F1-SCORE(micro) IS = 0.6438333333333334

Accuracy and F1-Scores for D value = 8

LSH ACCURACY IS = 66.06666666666666
LSH F1-SCORE(macro) IS = 0.5600191113776258
LSH F1-SCORE(micro) IS = 0.6606666666666666
PCA ACCURACY IS = 68.4
PCA F1-SCORE(macro) IS = 0.6378788719119015
PCA F1-SCORE(micro) IS = 0.684

Accuracy and F1-Scores for D value = 16
LSH ACCURACY IS = 65.46666666666667
LSH F1-SCORE(macro) IS = 0.5428425997058859
LSH F1-SCORE(micro) IS = 0.6546666666666666
PCA ACCURACY IS = 68.16666666666666
PCA F1-SCORE(macro) IS = 0.6322498661351897
PCA F1-SCORE(micro) IS = 0.6816666666666666

Accuracy and F1-Scores for D value = 32
LSH ACCURACY IS = 66.4
LSH F1-SCORE(macro) IS = 0.5599081366129793
LSH F1-SCORE(micro) IS = 0.664
PCA ACCURACY IS = 57.49999999999999
PCA F1-SCORE(macro) IS = 0.5149984606847011
PCA F1-SCORE(micro) IS = 0.575

Accuracy and F1-Scores for D value = 64
LSH ACCURACY IS = 65.3
LSH F1-SCORE(macro) IS = 0.5467126883449017
LSH F1-SCORE(micro) IS = 0.653
PCA ACCURACY IS = 63.73333333333333
PCA F1-SCORE(macro) IS = 0.5965566911415986
PCA F1-SCORE(micro) IS = 0.6373333333333333

Accuracy and F1-Scores for D value = 128
LSH ACCURACY IS = 65.51666666666667
LSH F1-SCORE(macro) IS = 0.535838581415689
LSH F1-SCORE(micro) IS = 0.6551666666666667
PCA ACCURACY IS = 71.36666666666667
PCA F1-SCORE(macro) IS = 0.6796538946366498
PCA F1-SCORE(micro) IS = 0.7136666666666667

Accuracy and F1-Scores for D value = 256
LSH ACCURACY IS = 66.55
LSH F1-SCORE(macro) IS = 0.567960491699557
LSH F1-SCORE(micro) IS = 0.6655
PCA ACCURACY IS = 72.26666666666667

PCA F1-SCORE(macro) IS = 0.6871978644283159
PCA F1-SCORE(micro) IS = 0.7226666666666667
Accuracy and F1-Scores for D value = 512
LSH ACCURACY IS = 65.51666666666667
LSH F1-SCORE(macro) IS = 0.5503758366301957
LSH F1-SCORE(micro) IS = 0.6551666666666667
PCA ACCURACY IS = 72.86666666666667
PCA F1-SCORE(macro) IS = 0.6927999919097676
PCA F1-SCORE(micro) IS = 0.7286666666666667

Accuracy and F1-Scores for D value = 1024
LSH ACCURACY IS = 65.28333333333333
LSH F1-SCORE(macro) IS = 0.5433573676573724
LSH F1-SCORE(micro) IS = 0.6528333333333334
PCA ACCURACY IS = 72.39999999999999
PCA F1-SCORE(macro) IS = 0.6865342213057751
PCA F1-SCORE(micro) IS = 0.724