

CMPT459 Assignment 4

Huiyi Zou 301355563

Question 1

1. For training data set:

Attribute A:

range: [3, 5], mean: 4.216, variance: 0.9533806684872497

Attribute B:

range: [0, 6], mean: 1.8584615384615384, variance: 0.6306065026169645

Attribute C:

range: [0, 3], mean: 1.549076923076923, variance: 0.2720634580619965

Attribute D:

range: [-2, 8], mean: -0.018307692307692306, variance: 1.264482693121333

Attribute E:

range: [-2, 8], mean: -0.13130769230769232, variance: 1.4293517265456956

Attribute F:

range: [-2, 8], mean: -0.16665384615384615, variance: 1.4198580285632287

Attribute G:

range: [-2, 8], mean: -0.21607692307692308, variance: 1.3701326914940517

Attribute H:

range: [-2, 8], mean: -0.2655769230769231, variance: 1.2794027211105754

Attribute I:

range: [-2, 8], mean: -0.28946153846153844, variance: 1.3151856787391232

Attribute J:

range: [4305, 16195], mean: 7281.844423076923, variance: 3565050.5029305494

Attribute K:

range: [10000, 800000], mean: 167693.83384615384, variance:
16790281030.879738

Attribute L:

range: [-165580, 746814], mean: 51499.240423076924, variance:
5400237087.585604

Attribute M:

range: [-69777, 743970], mean: 49416.01869230769, variance: 5039200503.96588

Attribute N:

range: [-157264, 1664089], mean: 47346.50796153846, variance:
4835932142.890971

Attribute O:

range: [-81334, 706864], mean: 43724.75861538461, variance:
4137279777.3071685

Attribute P:

range: [-81334, 823540], mean: 40532.02376923077, variance:
3687564920.1136703

Attribute Q:

range: [-209051, 568638], mean: 39097.61, variance: 3514823178.6675415

Attribute R:

range: [0, 873552], mean: 5671.37426923077, variance: 281485287.59756094

Attribute S:

range: [0, 1684259], mean: 5977.154846153846, variance: 569344638.9552133

Attribute T:

range: [0, 889043], mean: 5331.5725, variance: 302872455.0162444

Attribute U:

range: [0, 621000], mean: 4845.287076923077, variance: 248752249.31525284

Attribute V:

range: [0, 426529], mean: 4792.553846153846, variance: 232154508.44057906

Attribute W:

range: [0, 527143], mean: 5194.856961538461, variance: 300914573.9124213

For test data set:

Attribute A:

range: [3, 5], mean: 4.2956, variance: 0.912803200640128

Attribute B:

range: [0, 6], mean: 1.853, variance: 0.617114422884577

Attribute C:

range: [0, 3], mean: 1.541, variance: 0.2799749949989998

Attribute D:

range: [-2, 8], mean: -0.0424, variance: 1.2548532106421286

Attribute E:

range: [-2, 7], mean: -0.1814, variance: 1.4559852370474098

Attribute F:

range: [-2, 7], mean: -0.191, variance: 1.5084206841368273

Attribute G:

range: [-2, 7], mean: -0.2738, variance: 1.3623060212042406

Attribute H:

range: [-2, 7], mean: -0.3016, variance: 1.2944963392678537

Attribute I:

range: [-2, 8], mean: -0.3302, variance: 1.3510381676335266

Attribute J:

range: [4305, 15375], mean: 7378.688, variance: 3673011.8610282056

Attribute K:

range: [10000, 1000000], mean: 169096.0, variance: 17069476679.335867

Attribute L:

range: [-14386, 964511], mean: 49257.059, variance: 5554078780.027124

Attribute M:

range: [-24704, 983931], mean: 47296.0514, variance: 5156178732.923742

Attribute N:

range: [-17706, 578971], mean: 44721.1548, variance: 4701978117.727183

Attribute O:

range: [-170000, 891586], mean: 40650.7026, variance: 4154054826.2321997

Attribute P:

range: [-28335, 927171], mean: 38680.8152, variance: 3748447033.398128

Attribute Q:

range: [-339603, 961664], mean: 37196.3922, variance: 3692783593.4838758

Attribute R:

range: [0, 873552], mean: 5775.462, variance: 371424813.5604681

Attribute S:

range: [0, 1215471], mean: 5841.9232, variance: 581253297.4687955

Attribute T:

range: [0, 896040], mean: 5031.0452, variance: 528726275.205398

Attribute U:

range: [0, 621000], mean: 4713.3838, variance: 275194877.52160186

Attribute V:

range: [0, 332000], mean: 4797.295, variance: 235781858.65210542

Attribute W:

range: [0, 528666], mean: 5538.467, variance: 446865204.05092114

The total number of label 0: 20256

The total number of label 1: 5744

In this data set, there are 20256 tuples with class label 0, while there are only 5744 tuples with class label 1. The classes are not balanced.

2. [('D', 0.09504136618993976), ('E', 0.05347718601722567), ('F', 0.04176292436498385), ('G', 0.035058250631344545), ('H', 0.0319866717838871), ('I', 0.027124261451486758), ('A', 0.0021568158741513857), ('C', 0.0019010942704819182), ('B', 0.00114627975512127), ('J', 5.106433695522631e-11), ('R', 1.6358645786687293e-11), ('V', 1.4044812634307373e-11), ('U', 1.2031285108832723e-11), ('T', 1.1726975267559755e-11), ('W', 9.670443135633236e-12), ('S', 4.922007615953472e-12), ('K', 1.6465465934656703e-12), ('L', 8.714670425463054e-14), ('N', 5.0785362048150245e-14), ('M', 4.8128229859342215e-14), ('O', 3.676789745264193e-14), ('P', 1.6411518169588933e-14), ('Q', 1.2003584797033324e-14)]

Question 2

precision: 0.37723667220070095
recall: 0.3956277810021281
accuracy: 0.7222222222222222

precision: 0.38669399925456577
recall: 0.4037750535123565
accuracy: 0.7284188034188034

precision: 0.36228243315987796
recall: 0.39143078712679336
accuracy: 0.713974358974359

precision: 0.38677715726896056
recall: 0.4162466072120977
accuracy: 0.7258547008547008

precision: 0.365589519650655
recall: 0.4103117035875319
accuracy: 0.7162393162393162

precision: 0.3839205058717254
recall: 0.4132633216647219
accuracy: 0.7253418803418803

precision: 0.39924170616113747
recall: 0.405233788724264
accuracy: 0.7324786324786324

precision: 0.38680077642491617
recall: 0.4183206106870229
accuracy: 0.7212393162393163

precision: 0.39217056973086334
recall: 0.43013225991949394
accuracy: 0.7243162393162393

precision: 0.3954802259887006

recall: 0.41932367149758454

accuracy: 0.7298290598290599

Overall average precision: 0.3836193565712104

Overall average recall: 0.4103665584933995

Question 3

precision: 0.6929982046678635
recall: 0.2987038111820468
accuracy: 0.8158547008547009

precision: 0.6750889679715303
recall: 0.3691379645845495
accuracy: 0.8224358974358974

precision: 0.6618933969769292
recall: 0.32260566110895694
accuracy: 0.8143589743589743

precision: 0.6941786283891547
recall: 0.33753392787902287
accuracy: 0.8211965811965812

precision: 0.6790123456790124
recall: 0.3126837874926485
accuracy: 0.8179487179487179

precision: 0.6716772151898734
recall: 0.33022170361726955
accuracy: 0.8173504273504274

precision: 0.6706903055450774
recall: 0.3421204541081393
accuracy: 0.8165811965811965

precision: 0.6823529411764706
recall: 0.3320610687022901
accuracy: 0.8158119658119658

precision: 0.6691595441595442
recall: 0.3601686793176155
accuracy: 0.8176495726495726

precision: 0.65900107797341
recall: 0.35439613526570046
accuracy: 0.8166666666666667

Overall average precision: 0.6756052627728866
Overall average recall: 0.3359633193258239

The precision and the accuracy have been improved after using AdaBoost, while the value of recall is decreasing.

Question 4

For this question, I apply MLP classification.