Derya TINMAZ 2380947

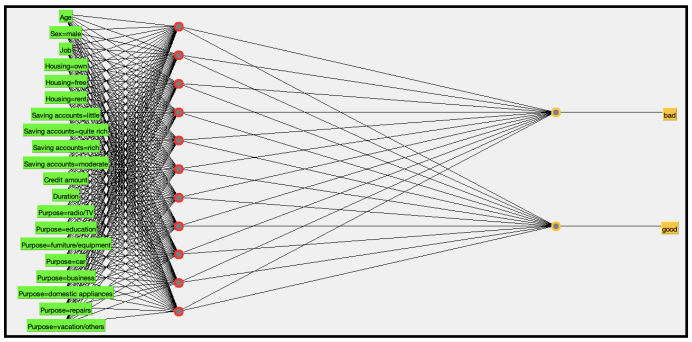
CENG 414 Introduction to Data Mining

Spring 2023 - Assignment 2

2.2 Multi-Layer Perceptron

Default Paramaters:

- learning rate: 0.3
- momentum: 0.2
- number of epochs: 500
- percentage size of validation set: 0
- seed: 0
- threshold for number of consecutive errors: 20
- comma separated numbers for nodes on each layer : a ((attribs + classes) / 2)



1. There is one hidden layer, 11 hidden nodes are created.

2. By default Weka normalizes the attributes (rescales the attributes to the range of 0 to 1) . The effect of the normalization is that it can help training of the neural networks as the different features are on a similar scale, which helps to stabilize the gradient descent step, allowing us to use larger learning rates or help models converge faster for a given learning rate.

3. It uses early stopping method with the validation threshold parameter, which set equal to 20 by default. WEKA will stop the learning if the number of consecutive error that occur is more than 20.

Detailed Accuracy By Class and Confusion Matrix:

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.313    0.173    0.469      0.313    0.376      0.159    0.616     0.417     bad
                 0.827    0.687    0.712      0.827    0.765      0.159    0.616     0.753     good
Weighted Avg.    0.659    0.518    0.632      0.659    0.637      0.159    0.616     0.643

=== Confusion Matrix ===

   a   b   <-- classified as
  84 184 |   a = bad
  95 454 |   b = good
```

accuracy = (84 + 454) / (84+184+95+454) = 0.659
precision = 454 / (454 + 184) = 0.712
recall = 454 / (454 + 95) =. 0.827
F1-measure = 2*454 / (2*454 + 95 + 184) = 0.765

2.3 Decision Tree

Pruned Tree:

```
J48 pruned tree
------------------

Duration <= 42
|   Credit amount <= 7814
|   |   Housing = own: good (533.0/133.0)
|   |   Housing = free: good (60.0/20.0)
|   |   Housing = rent
|   |   |   Job <= 2
|   |   |   |   Duration <= 15: good (67.0/19.0)
|   |   |   |   Duration > 15
|   |   |   |   |   Age <= 45: bad (49.0/19.0)
|   |   |   |   |   Age > 45: good (5.0/1.0)
|   |   |   Job > 2: good (15.0/2.0)
|   Credit amount > 7814
|   |   Duration <= 36: bad (32.0/8.0)
|   |   Duration > 36: good (4.0)
Duration > 42: bad (52.0/13.0)

Number of Leaves  :     9

Size of the tree :      16
```

Summary :

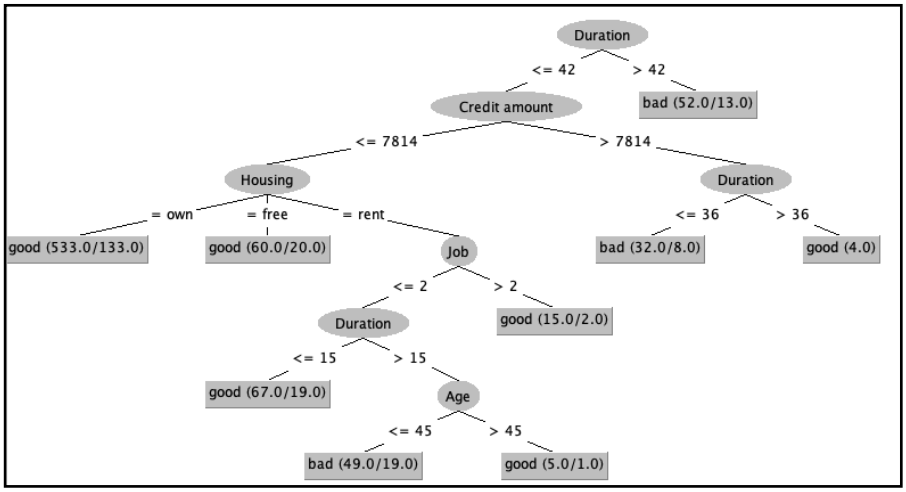
```
=== Summary ===

Correctly Classified Instances         532               65.1163 %
Incorrectly Classified Instances       285               34.8837 %
Kappa statistic                          0.1149
Mean absolute error                      0.4127
Root mean squared error                  0.5044
Relative absolute error                 93.5663 %
Root relative squared error            107.4262 %
Total Number of Instances              817
```

Detailed Accuracy By Class and Confusion Matrix:

```
=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.261    0.158    0.446      0.261    0.329      0.122    0.585     0.385     bad
               0.842    0.739    0.700      0.842    0.764      0.122    0.585     0.728     good
Weighted Avg.  0.651    0.548    0.617      0.651    0.622      0.122    0.585     0.616

=== Confusion Matrix ===

   a   b   <-- classified as
  70 198 |   a = bad
  87 462 |   b = good
```

accuracy = (70 + 462) / (70+198+87+462) = 0.651
precision = 462 / (462 + 198) = 0.7
recall = 462 / (462 + 87) = 0.842
F1-measure = 2*462 / (2*462 + 87 + 198) = 0.764

Visualization of the Tree:



2.4 Naive Bayes

Detailed Accuracy By Class and Confusion Matrix:

```
=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.276    0.107    0.556      0.276    0.369      0.214    0.676     0.514     bad
                0.893    0.724    0.716      0.893    0.795      0.214    0.676     0.792     good
Weighted Avg.   0.690    0.522    0.664      0.690    0.655      0.214    0.676     0.701

=== Confusion Matrix ===

   a   b   <-- classified as
  74 194 |   a = bad
  59 490 |   b = good
```

accuracy = (74 + 490) / (74+194+59+490) = 0.69
precision = 490 / (490 + 194) = 0.716
recall = 490 / (490 + 59) = 0.893
F1-measure = 2*490 / (2*490 + 59 + 194) = 0.795

Classifier Model:

```
Naive Bayes Classifier

                          Class
Attribute                   bad      good
                          (0.33)    (0.67)
=============================================
Age
  mean                  33.8307   35.6934
  std. dev.              11.339   11.3467
  weight sum                268       549
  precision              1.0769    1.0769

Sex
  female                  97.0     166.0
  male                   173.0     385.0
  [total]                270.0     551.0

Job
  mean                   1.9291     1.878
  std. dev.              0.6682    0.6456
  weight sum                268       549
  precision                   1         1

Housing
  own                    166.0     419.0
  free                    40.0      47.0
  rent                    65.0      86.0
  [total]                271.0     552.0
```

```
Saving accounts
  little                218.0     387.0
  quite rich             12.0      53.0
  rich                    7.0      43.0
  moderate               35.0      70.0
  [total]               272.0     553.0

Credit amount
  mean                3887.287 2758.8841
  std. dev.           3491.794 2173.2034
  weight sum               268       549
  precision              23.8817   23.8817

Duration
  mean                  25.0718   18.2115
  std. dev.             13.7023   10.0094
  weight sum                268       549
  precision               2.125     2.125

Purpose
  radio/TV                57.0     173.0
  education               23.0      23.0
  furniture/equipment     55.0     107.0
  car                     90.0     175.0
  business                33.0      52.0
  domestic appliances      4.0       7.0
  repairs                  9.0      13.0
  vacation/others          5.0       7.0
  [total]                276.0     557.0
```