CENG 414
Introduction to Data Mining
Fall 2019-2020
THE 2

Question 5

1- Report summary and detailed accuracy by class.

```
=== Summary ===

Correctly Classified Instances         211               73.2639 %
Incorrectly Classified Instances        77               26.7361 %
Kappa statistic                          0.6461
Mean absolute error                      0.2812
Root mean squared error                  0.3572
Relative absolute error                 74.9451 %
Root relative squared error             82.3928 %
Total Number of Instances              288

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.758    0.208    0.500      0.758    0.603      0.482    0.776     0.431     opel
                 0.346    0.038    0.771      0.346    0.478      0.419    0.788     0.539     saab
                 0.870    0.028    0.918      0.870    0.893      0.856    0.941     0.856     bus
                 0.986    0.074    0.814      0.986    0.892      0.859    0.962     0.811     van
Weighted Avg.    0.733    0.081    0.763      0.733    0.718      0.658    0.869     0.668
```

2- Explain the C parameter of SVM. Change this parameter and run the classifier with various values: Plot and report the effects.

By official javadoc of weka, C parameter is

```
C <double>
  The complexity constant C. (default 1)
```

SMO function is stands for sequential minimal optimization and this algorithm uses a support vector machine for training. Goals of this algorithm are classifyiing data in a better way and minimizing the misclassification of instances. When the algorithm becomes more greedy to get the correct classification(first goal), it's misclassification rate getting higher as well. Also, when we try to minimize misclassification, model will get less result.

C parameter exists due to these conflicting goals. If C value becomes larger, smaller-margin hyperplanes are chosen and model avoids misclassification more. When C value becomes smaller, much more misclassified instances occur.

Thus, C is the parameter that tells the program how much it should avoid from misclassifying.

When c=2,

```
=== Summary ===

Correctly Classified Instances         216               75      %
Incorrectly Classified Instances        72               25      %
Kappa statistic                          0.6682
Mean absolute error                      0.2781
Root mean squared error                  0.353
Relative absolute error                 74.0969 %
Root relative squared error             81.4218 %
Total Number of Instances              288

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.694     0.186     0.506      0.694     0.585      0.458    0.766     0.440     opel
               0.449     0.057     0.745      0.449     0.560      0.471    0.783     0.543     saab
               0.883     0.024     0.932      0.883     0.907      0.874    0.956     0.885     bus
               0.986     0.060     0.843      0.986     0.909      0.881    0.969     0.841     van
Weighted Avg.  0.750     0.077     0.768      0.750     0.744      0.677    0.872     0.686
```

And when c=3,

```
=== Summary ===

Correctly Classified Instances         218               75.6944 %
Incorrectly Classified Instances        70               24.3056 %
Kappa statistic                          0.6773
Mean absolute error                      0.2772
Root mean squared error                  0.3519
Relative absolute error                 73.8656 %
Root relative squared error             81.1693 %
Total Number of Instances              288

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.710     0.186     0.512      0.710     0.595      0.470    0.774     0.445     opel
               0.462     0.067     0.720      0.462     0.563      0.463    0.774     0.534     saab
               0.883     0.024     0.932      0.883     0.907      0.874    0.957     0.886     bus
               0.986     0.041     0.886      0.986     0.933      0.912    0.978     0.884     van
Weighted Avg.  0.757     0.075     0.773      0.757     0.753      0.685    0.873     0.695
```

And lastly when c= 10,

```
=== Summary ===

Correctly Classified Instances         226               78.4722 %
Incorrectly Classified Instances        62               21.5278 %
Kappa statistic                          0.7135
Mean absolute error                      0.2737
Root mean squared error                  0.3469
Relative absolute error                 72.9404 %
Root relative squared error             80.0234 %
Total Number of Instances              288

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.694    0.155    0.551      0.694    0.614      0.498    0.784     0.470     opel
                 0.564    0.086    0.710      0.564    0.629      0.517    0.798     0.568     saab
                 0.896    0.019    0.945      0.896    0.920      0.892    0.960     0.902     bus
                 0.986    0.023    0.933      0.986    0.959      0.946    0.988     0.932     van
Weighted Avg.    0.785    0.067    0.794      0.785    0.785      0.719    0.885     0.726
```

Effect of c parameter is, when it is increased, the algorithm gets more greedy to classify more elements to it's correct class.

3- Explain the terms maximum margin hyperplane and support vector.

Hyperplane -> A separator between 2 class of data.

Maximum-margin hyperplane -> A separator that maximizes distance between the hyperplane and training examples which consist of two different categories.

Support Vectors -> The vectors that define the hyperplanes which is on the closest training examples to maximum margin hyperplane.

4- When we run SVM in Weka, it uses a kernel function in default. What is a kernel function? What is the benefit to use a kernel function? Explain clearly

This Kernel function known as kernel trick as well in machine learning. Kernel functions are used to transform nearly inseparable data to linearly separable ones. This function is applied on every data instance. These data instances are mapped to higher dimensional space and they become linearly seperable.

Alper KOCAMAN - 2169589