

Luiza Ávila

1)

A etapa que eu apliquei na base de dados “**contact-lenses.arff**” foi a de balanceamento, pois há uma classe bem maior que as outras, desbalanceando os dados. Apliquei ambos os filtros “Resample” e “SpreadSubsample” fornecidos pelo weka, que produzem uma sub amostra aleatória de um conjunto de dados, com o objetivo de equilibrar de maneira mais justa as classes dos dados. Fazendo uma sub amostragem aleatória, consigo eliminar exemplos de forma randômica, equilibrando de forma mais justa as duas classes (maioritária e minoritária), afim de evitar problemas de oversample ou undersample.

```
=== Summary ===

Correctly Classified Instances      21           87.5   %
Incorrectly Classified Instances    3           12.5   %
Kappa statistic                    0.7592
Mean absolute error                 0.1238
Root mean squared error             0.2558
Relative absolute error             32.7657 %
Root relative squared error        58.5762 %
Total Number of Instances          24

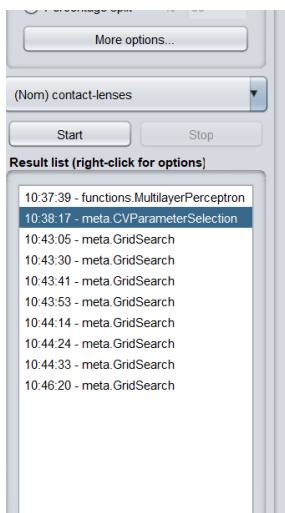
=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC       ROC Area  PRC Area  Class
          1,000    0,000    1,000     1,000    1,000      1,000     1,000     1,000     soft
          0,500    0,050    0,667     0,500    0,571      0,507     0,875     0,570     hard
          0,933    0,222    0,875     0,933    0,903      0,730     0,941     0,974     none
Weighted Avg.   0,875    0,147    0,866     0,875    0,868      0,749     0,942     0,912

=== Confusion Matrix ===

 a  b  c   <-- classified as
 5  0  0 | a = soft
 0  2  2 | b = hard
 0  1 14 | c = none
```

2) Eu apliquei o CVParameter Selection do weka. Nada foi alterado em relação aos resultados anteriores. Tentei aplicar também o grid search, mas ele não funciona com o mlps.



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      21           87.5   %
Incorrectly Classified Instances    3           12.5   %
Kappa statistic                    0.7592
Mean absolute error                 0.1238
Root mean squared error             0.2558
Relative absolute error             32.7657 %
Root relative squared error        58.5762 %
Total Number of Instances          24

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC       ROC Area  PRC Area  Class
          1,000    0,000    1,000     1,000    1,000      1,000     1,000     1,000     soft
          0,500    0,050    0,667     0,500    0,571      0,507     0,875     0,570     hard
          0,933    0,222    0,875     0,933    0,903      0,730     0,941     0,974     none
Weighted Avg.   0,875    0,147    0,866     0,875    0,868      0,749     0,942     0,912

=== Confusion Matrix ===

 a  b  c   <-- classified as
 5  0  0 | a = soft
 0  2  2 | b = hard
 0  1 14 | c = none
```