Overall analysis of one data set and different performance measures

R supplement of

"Exploratory and Inferential Analysis of Benchmark Experiments"

Manuel J. A. Eugster

benchmark version 0.01

In Section 3.3 – Overall analysis – we use ranking and order mechanisms to sum up the different orders which the exploratory and inferential analyses resulted in.

Requirements:

```
> Sweave('uci621.Rnw')
> Sweave('uci621-1x1-exp.Rnw')
> Sweave('uci621-1x1-inf.Rnw')
```

We want to evaluate the candidate algorithms according to their worst case misclassification performance, the relatively misclassification performance to each other, and their mean computation time.

For the worst case we have calculated the minimax ranking:

```
> minimax.q95
```

```
       svm      rpart        lda naiveBayes       nnet        knn
         1          2          3          3          5          6
```

For the relatively performance to each other we have calculated a relation based on the pairwise tests based on a mixed effects model:

```
> rel
```

```
A binary relation of size 6 x 6.
```

```
> tsort(rel)
```

```
 rpart - svm < nnet < knn - lda - naiveBayes
```

For the mean computation time we can calculate a simple ranking:

```
> mtime <- apply(uci621[,,'Time','monks3'], 'alg', mean)
```

```
       lda naiveBayes        knn      rpart        svm       nnet
   0.07260    0.34876    2.56056    0.05516   12.22100    9.22524
```

```
> time <- as.ranking(mtime)
```

```
     rpart        lda naiveBayes        knn       nnet        svm
         1          2          3          4          5          6
```

1 Consensus

The consensus is an aggregation of preferences of voters where all are equally relevant. To calculate the consensus all ranking and order information has to be adapted as relations.

```
> rw <- as.relation(minimax.q95)
> rm <- rel
> rc <- as.relation(time)
```

The consensus is built upon a class of relations, e.g., linear orders.

```
> con1 <- relation_consensus(relation_ensemble(rw, rm, rc), 'SD/L',
+                            control=list(all=TRUE))
```

The consensus does not have to be unique; in this case the consensus consists of two relations.

```
> tsort(con1)
```

```
[[1]]
 svm < rpart < lda < naiveBayes < nnet < knn

[[2]]
 rpart < svm < lda < naiveBayes < nnet < knn
```

We decide that in case of our problem the worst case performance is more important then the other performance measures we can weight the relations.

```
> con2 <- relation_consensus(relation_ensemble(rw, rm, rc), 'SD/L',
+                            control=list(all=TRUE),
+                            weights=c(1.2, 1, 1))
> tsort(con2)
```

```
[[1]]
 svm < rpart < lda < naiveBayes < nnet < knn
```

References

M. J. A. Eugster, T. Hothorn, and F. Leisch. Exploratory and inferential analysis of benchmark experiments. Technical Report 30, Institut für Statistik, Ludwig-Maximilians-Universität München, Germany, 2008. URL http://epub.ub.uni-muenchen.de/4134/.