

Inferential analysis of one data set and one performance measure

R supplement of

“Exploratory and Inferential Analysis
of Benchmark Experiments”

Manuel J. A. Eugster

benchmark version 0.01

In Section 3.2 – Inferential analysis – we use more formal tools to infer a statistically correct order of the candidate algorithms.

Requirements:

```
> Sweave('uci621.Rnw')
```

1 A Non-parametric analysis

The non-parametric analysis is based upon Friedman-rank based tests.

```
> ibea <- make.fr.ibea()
```

Global test: As global test we use the Friedman test.

```
> gt <- ibea$test.global(monks3)
```

Asymptotic Friedman Test

```
data: value by  
      alg (lda, naiveBayes, knn, rpart, svm, nnet)  
      stratified by samp  
chi-squared = 888.5657, df = 5, p-value < 2.2e-16
```

The specific measures are

```
> statistic(gt, type='linear')
```

```
lda          1175.5  
naiveBayes  1176.5  
knn          1135.5  
rpart        413.5
```

```
svm          393.5
nnet         955.5
```

```
> pvalue(gt)
```

```
[1] 0
```

Pairwise tests: As pairwise test we use the Wilcoxon-Nemenyi-McDonald-Thompson test.

```
> pt <- ibea$test.pairwise(monks3)
```

```
Asymptotic General Independence Test
```

```
data: value by
      alg (lda, naiveBayes, knn, rpart, svm, nnet)
      stratified by samp
maxT = 19.721, p-value < 2.2e-16
```

```
> cbind(statistic(pt, type='linear'),
+       pvalue(pt, method='single-step'))
```

```
naiveBayes - lda      1 1.000000e+00
knn - lda            -40 9.156536e-01
rpart - lda          -762 0.000000e+00
svm - lda            -782 0.000000e+00
nnet - lda           -220 9.034885e-07
knn - naiveBayes     -41 9.070688e-01
rpart - naiveBayes  -763 0.000000e+00
svm - naiveBayes     -783 0.000000e+00
nnet - naiveBayes   -221 3.602279e-07
rpart - knn          -722 0.000000e+00
svm - knn            -742 0.000000e+00
nnet - knn           -180 9.033753e-05
svm - rpart          -20 9.960508e-01
nnet - rpart         542 0.000000e+00
nnet - svm           562 0.000000e+00
```

Relation and algorithm order: According to the significance value 0.05 we can establish a relation based on the pairwise tests.

```
> rel <- ibea$relation.pairwise(pt, 0.05)
```

A binary relation of size 6 x 6.

The relation has the domain

```
> relation_domain(rel)
```

Relation domain:

A pair with elements:

```
{knn, lda, naiveBayes, nnet, rpart, svm}
{knn, lda, naiveBayes, nnet, rpart, svm}
```

and the graph

```
> relation_graph(rel)
```

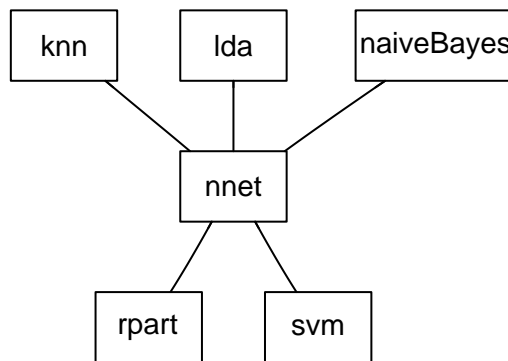
Relation graph:

A set with pairs:

```
(nnet, knn)
(nnet, lda)
(nnet, naiveBayes)
(rpart, knn)
(rpart, lda)
(rpart, naiveBayes)
(rpart, nnet)
(svm, knn)
(svm, lda)
(svm, naiveBayes)
(svm, nnet)
```

A Hasse diagram visualizes the relation.

```
> plot(rel)
```



A topological sort algorithm brings the relation in a order (non-uniquess indicates non-significance, in the article shown with \approx in the R output shown with -).

```
> ord <- tsort(rel)
```

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

```
> as.ranking(ord)
```

| rpart | svm | nnet | knn | lda | naiveBayes |
|-------|-----|------|-----|-----|------------|
| 1 | 1 | 3 | 4 | 4 | 4 |

2 A Parametric analysis

The parametric analysis approach is based upon linear mixed effects models.

```
> ibea <- make.lmer.ibea()
```

Model: First step is fitting a linear mixed effects model:

```
> model <- ibea$model(monks3)
```

Linear mixed model fit by REML

Formula: value ~ alg + (1 | samp)

Data: data

| AIC | BIC | logLik | deviance | REMLdev |
|-------|-------|--------|----------|---------|
| -9756 | -9713 | 4886 | -9850 | -9772 |

Random effects:

| Groups | Name | Variance | Std.Dev. |
|--------|-------------|------------|-----------|
| samp | (Intercept) | 2.6592e-05 | 0.0051567 |
| | Residual | 6.7544e-05 | 0.0082185 |

Number of obs: 1500, groups: samp, 250

Fixed effects:

| | Estimate | Std. Error | t value |
|----------------|------------|------------|---------|
| (Intercept) | 3.522e-02 | 6.136e-04 | 57.40 |
| algnaiiveBayes | 4.188e-05 | 7.351e-04 | 0.06 |
| algknn | -7.783e-04 | 7.351e-04 | -1.06 |
| algrpart | -2.359e-02 | 7.351e-04 | -32.09 |
| algsvm | -2.427e-02 | 7.351e-04 | -33.02 |
| algnnet | -5.899e-03 | 7.351e-04 | -8.03 |

Correlation of Fixed Effects:

| | (Intr) | algnvB | algknn | algrpr | algsvm |
|--------------|--------|--------|--------|--------|--------|
| algnaiiveBys | -0.599 | | | | |
| algknn | -0.599 | 0.500 | | | |
| algrpart | -0.599 | 0.500 | 0.500 | | |
| algsvm | -0.599 | 0.500 | 0.500 | 0.500 | |
| algnnet | -0.599 | 0.500 | 0.500 | 0.500 | 0.500 |

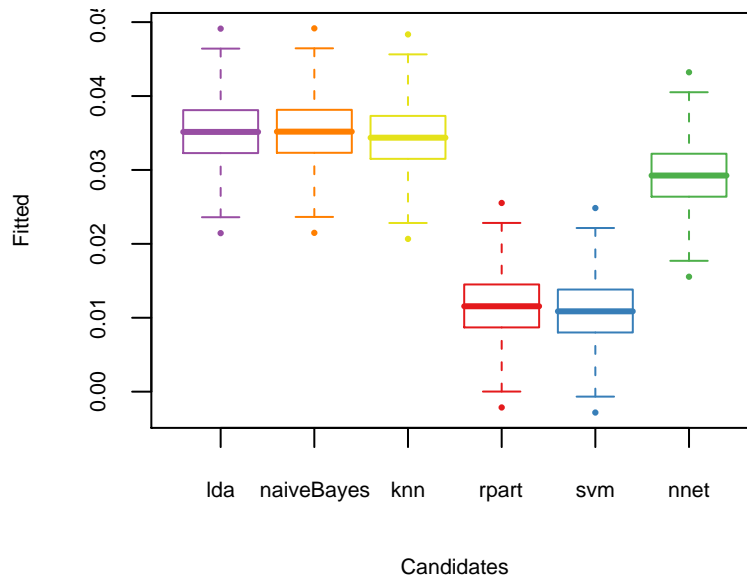
We can apply different diagnostic plots to check the quality of the model.

Fitted values plot:

```

> par(mar=c(4,4,0,0)+0.1, ps=8)
> boxplot(fitted(model) ~ alg, data=melt(monks3),
+         border=sixcols, pch=19, cex=0.3,
+         xlab='Candidates', ylab='Fitted')

```

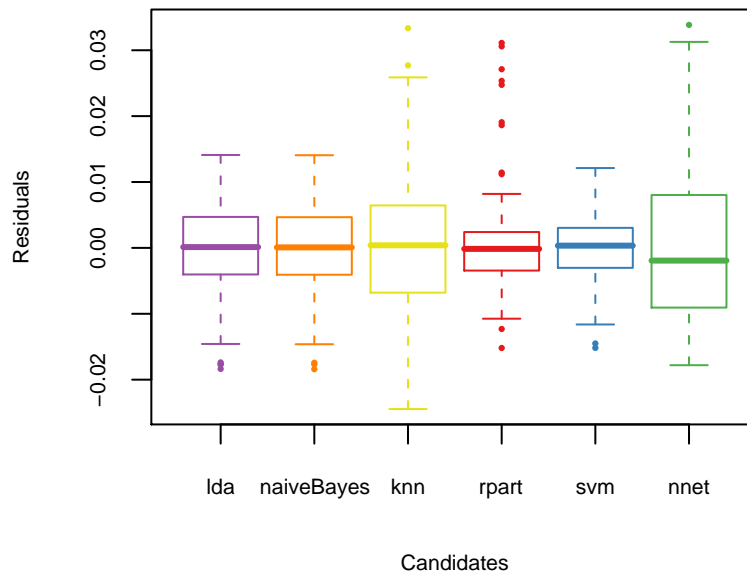


Residual plot:

```

> par(mar=c(4,4,0,0)+0.1, ps=8)
> boxplot(resid(model) ~ alg, data=melt(monks3),
+         border=sixcols, pch=19, cex=0.3,
+         xlab='Candidates', ylab='Residuals')

```



Pairwise tests: Based on the fitted model we test for pairwise differences between the algorithms.

```

> pt <- ibea$test.pairwise(model)

```

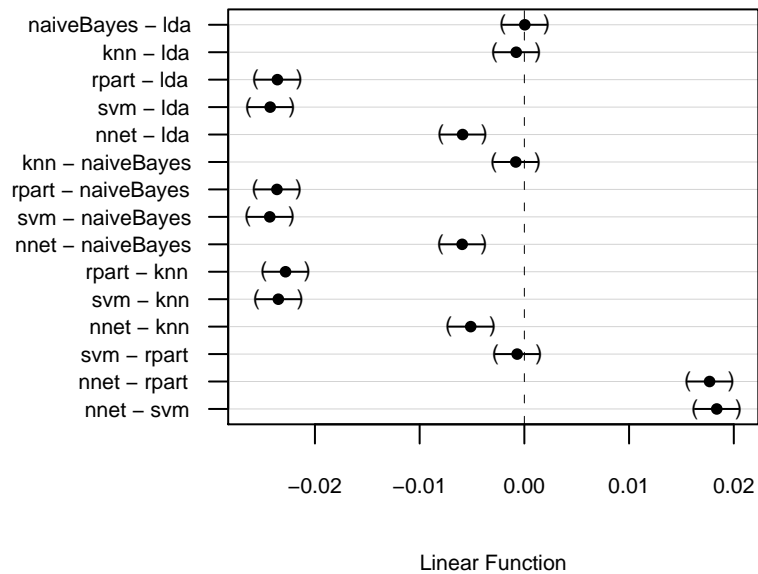
General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Linear Hypotheses:

| | Estimate |
|-------------------------|------------|
| naiveBayes - lda == 0 | 4.188e-05 |
| knn - lda == 0 | -7.783e-04 |
| rpart - lda == 0 | -2.359e-02 |
| svm - lda == 0 | -2.427e-02 |
| nnet - lda == 0 | -5.899e-03 |
| knn - naiveBayes == 0 | -8.202e-04 |
| rpart - naiveBayes == 0 | -2.363e-02 |
| svm - naiveBayes == 0 | -2.431e-02 |
| nnet - naiveBayes == 0 | -5.941e-03 |
| rpart - knn == 0 | -2.281e-02 |
| svm - knn == 0 | -2.349e-02 |
| nnet - knn == 0 | -5.121e-03 |
| svm - rpart == 0 | -6.863e-04 |
| nnet - rpart == 0 | 1.769e-02 |
| nnet - svm == 0 | 1.837e-02 |

```
> par(mar=c(4,6,0,0)+0.1, ps=8)
> plot(pt)
```

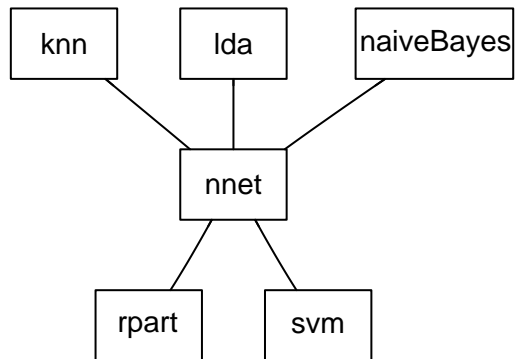


Relation and algorithm order: As from this point everything is similar to the non-parametric approach.

```
> rel <- ibea$relation.pairwise(pt, 0.05)
```

A binary relation of size 6 x 6.

```
> plot(rel)
```



```
> ord <- tsort(rel)
```

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

```
> as.ranking(ord)
```

| rpart | svm | nnet | knn | lda | naiveBayes |
|-------|-----|------|-----|-----|------------|
| 1 | 1 | 3 | 4 | 4 | 4 |

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/>.