Example_CVlm.R

gerkovink

2020-01-03

load the required package

```
library(DAAG)
```

```
## Loading required package: lattice
```

transformed variables (log-transformed)

```
CVlm(data=nihills, form.lm = formula(log(time) ~ log(climb) + log(dist)),
     plotit="Observed")
```

example from the CVlm() function

```
## Analysis of Variance Table
##
## Response: log(time)
##             Df Sum Sq Mean Sq F value  Pr(>F)
## log(climb)   1   5.94    5.94    1013 < 2e-16 ***
## log(dist)    1   0.89    0.89     152 8.2e-11 ***
## Residuals   20   0.12    0.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

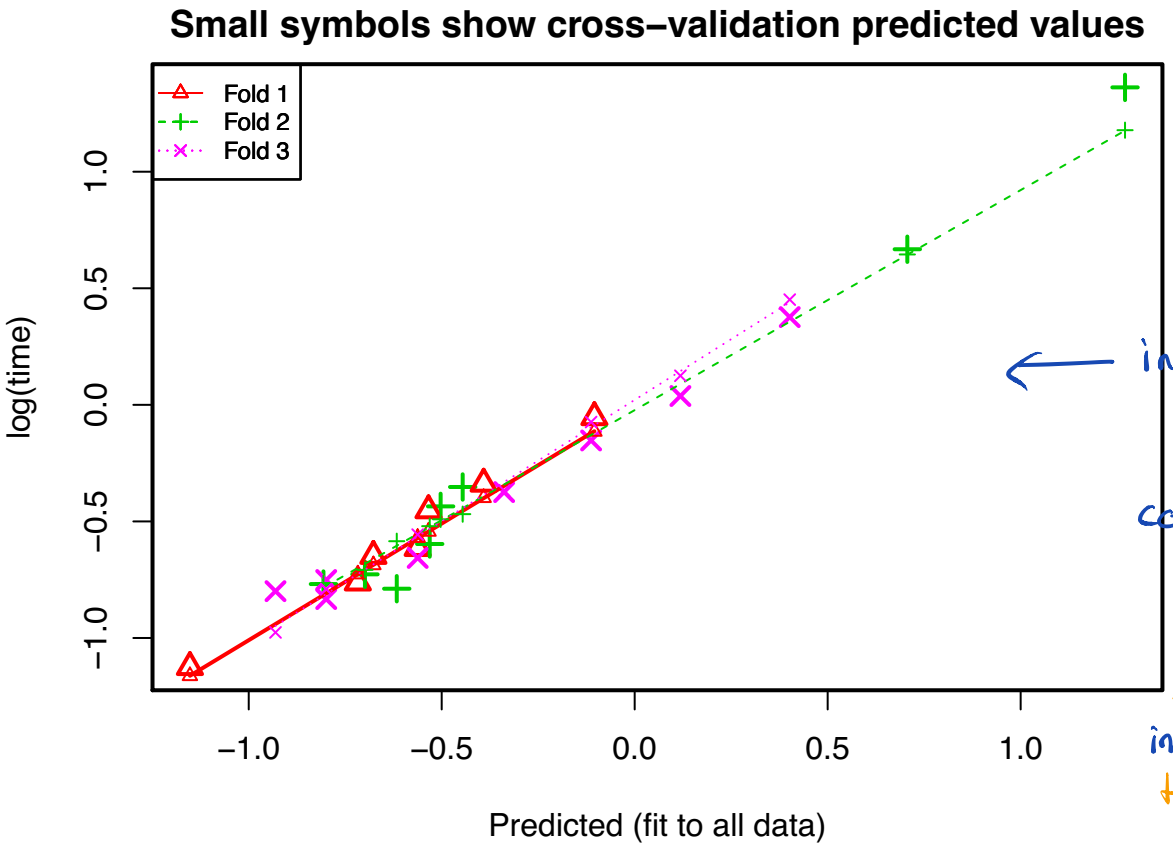
standard ANOVA table

residual variance for complete data → Sum Sq / Df

```
## Warning in CVlm(data = nihills, form.lm = formula(log(time) ~ log(climb) + :
##
##  As there is >1 explanatory variable, cross-validation
##  predicted values for a fold are not a linear function
##  of corresponding overall predicted values.  Lines that
##  are shown for the different folds are approximate
```

just a warning that you have more than one predictor and that the resulting crossvalidation is thus based on a set of additive effects.

log-transformation makes more like



← inspection:
lines seem very similar
conclusion:
no important visual difference between the three folds
implication:
the model seems to fit more-or-less equally to the three folds.
Hence, evidence for a useful model

```
##
## fold 1
## Observations in test set: 7
##              Slieve Gullion McVeigh Classic   Rocky Slieve Donard
## Predicted           -0.7174         -0.5623 -0.6770       -0.1044
## cvpred              -0.7264         -0.5702 -0.6870       -0.1147
## log(time)           -0.7621         -0.6141 -0.6481       -0.0530
## CV residual         -0.0358         -0.0439  0.0389        0.0616
##              Scrabo Hill Race Slieve Gallion BARF Turkey Trot
## Predicted             -1.1523         -0.534          -0.3913
## cvpred                -1.1627         -0.543          -0.3989
## log(time)             -1.1248         -0.452          -0.3382
## CV residual            0.0379          0.091           0.0607
##
## Sum of squares = 0.02    Mean square = 0    n = 7
##
## fold 2
## Observations in test set: 8
##              Glenariff Mountain Tollymore Mountain Slieve Martin
## Predicted                -0.446            -0.6989       -0.5310
## cvpred                   -0.469            -0.7078       -0.5204
## log(time)                -0.352            -0.7270       -0.5968
## CV residual               0.117            -0.0192       -0.0764
##              Moughanmore Annalong Horseshoe Loughshannagh Horseshoe
## Predicted        -0.80639              0.7063                -0.5028
## cvpred           -0.76424              0.6451                -0.4910
## log(time)        -0.76871              0.6674                -0.4355
```

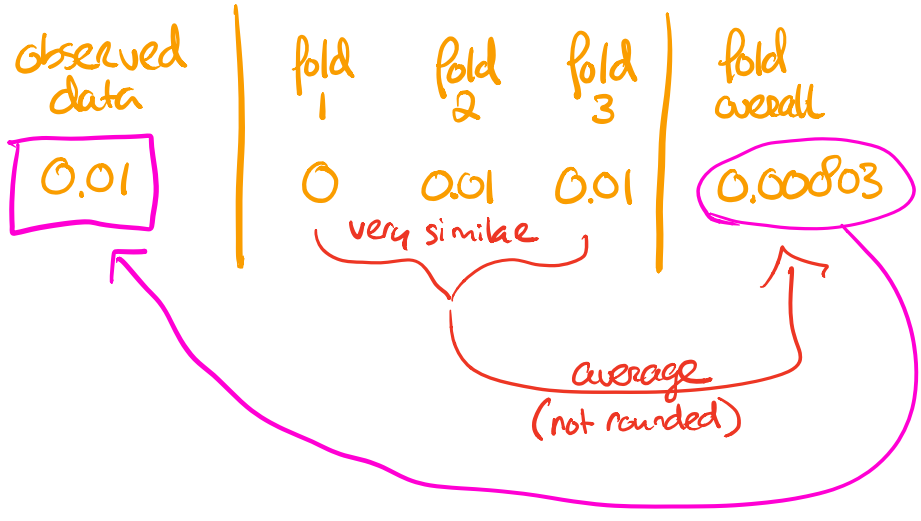
residual variance for fold # 1

```
## CV residual    -0.00447             0.0223                  0.0555
##             Meelbeg Meelmore Seven Sevens
## Predicted             -0.616        1.270
## cvpred                -0.585        1.178
## log(time)             -0.789        1.362
## CV residual           -0.204        0.184
##
## Sum of squares = 0.1    Mean square = 0.01    n = 8
##
## fold 3
## Observations in test set: 8
##             Binevenagh Donard & Commedagh Hen & Cock Monument Race
## Predicted      -0.1129             0.1184     -0.931        -0.799
## cvpred         -0.0738             0.1249     -0.976        -0.790
## log(time)      -0.1528             0.0379     -0.799        -0.751
## CV residual    -0.0789            -0.0870      0.177         0.039
##             Donard Forest Flagstaff to Carling Slieve Bearnagh
## Predicted         -0.5623                0.402        -0.33834
## cvpred            -0.5568                0.451        -0.36764
## log(time)         -0.6566                0.376        -0.37429
## CV residual       -0.0998               -0.075        -0.00665
##             Lurig Challenge
## Predicted           -0.7992
## cvpred              -0.7905
## log(time)           -0.8330
## CV residual         -0.0426
##
## Sum of squares = 0.06    Mean square = 0.01    n = 8
##
## Overall (Sum over all 8 folds)
##       ms
## 0.00803
```

residual variance for fold #2

residual variance for fold #3

overall residual variance for cross validation



very similar!
Here more evidence for a useful, well-fitted model that can be generalized to the population