Assessing model quality while deploying logistic regression

Gonçalo Hora de Carvalho

November 2018

1. Logistic Regression

We asked weather it is possible to predict if participant 8 is going to get a particular trial correct, that is to say, if the variable ER can be accurately predicted. To determine this a logistic regression of ER was ran on RT, blocknum, isDots, cohFac, and isLeft.

It was found that some independent variables are connected to the trial outcome in a statistical significant way, that is, to the dependent variable of a particular trial being correct for participant number 8. These variables are: 'RT', 'isDots', 'cohFac' and 'isleft'. Their means or effects were then verified to be statistically relevant by deploying an ANOVA test. The most informative variable in predicting the participants trial outcome seems to be the 'RT' variable - reaction times. An interpretation is that response time correlates to interpretation time - the faster the participant computes a visual understanding of the game, leading to a visual answer (the faster the player interprets the game) the faster s/he will answer. The factor 'isDots' has been found to have the 2nd most impact in a significant positive effect on the outcome of the trial. Perhaps due to a simplification of the test being related to a better performance. The variable 'isLeft' also bares significant weight on the results. This might correlate to the fact that the author is right handed vs being left handed - some studies have found that visual directionality input might correlate to hand-preference task (insert citation). The variable 'cohFac' - the proportion of dots that move into a single direction, has a significant negative effect on the ER - since the cohFac acts as a measurement of task difficulty this is expect to be the case.

2. Model Assessment

In fig.1 the plot of the receiver operating characteristic (ROC) curve can be analysed. The data forms an increasing over arching curve indicating an increase of sensitivity over specificity through the domain. The test can be said to have been accurate since the curve is very close to the left-hand border (closer to the true positive rate values) as well as to the top of the ROC space. The area underneath the curve can be used as a measurement of test accuracy (the larger the area the more accurate the test is) - which for the purpose of the test in question points towards its accuracy.

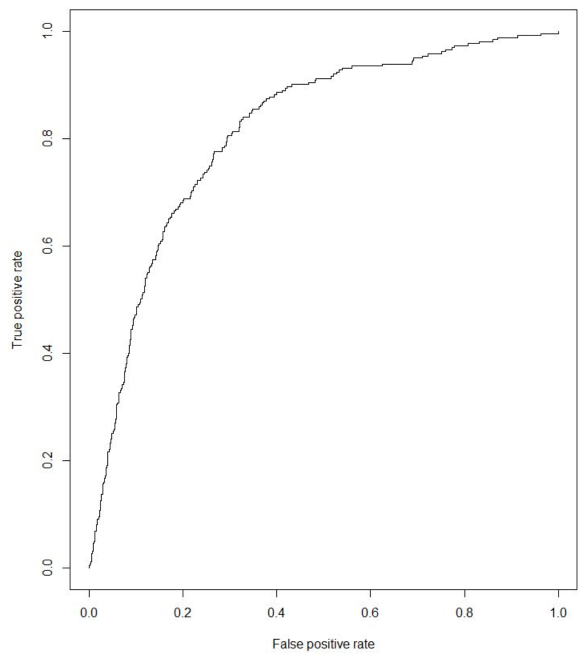

fig.1 - Performance Plot

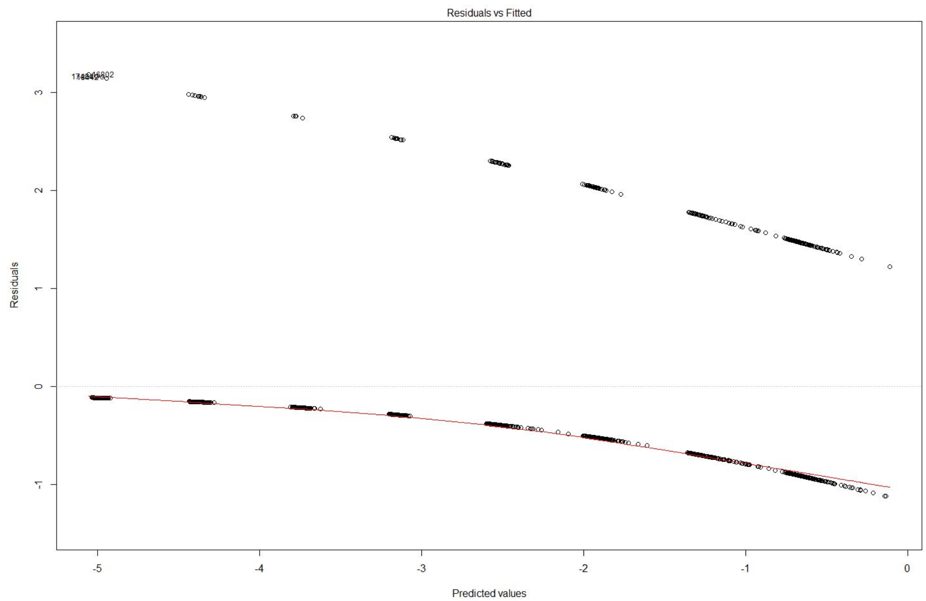

Fig.2 - Residuals vs Fitted plot.

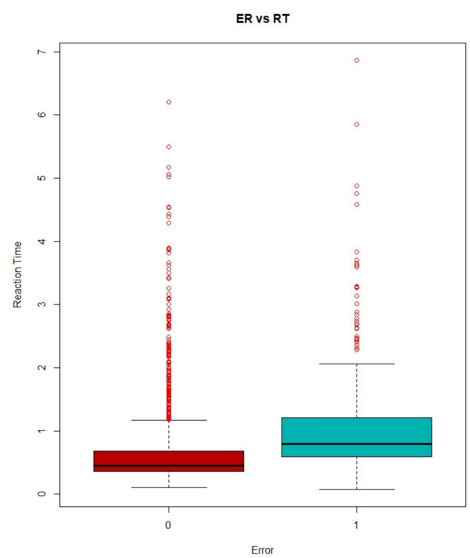

Fig.3 - Boxplot ER vs RT.

An analyses over the plot in fig. 2 reveals that the data seems to be skewed, with a heavy tail on the right of the plot. This is unexpected, since the data should follow a normal distribution. Such large RT's seem to be too large to have occurred naturally. If it is the case that they occurred for some particular reason, that is, if they weren't simply due to the participant's slow reaction, then these outliers can and should be removed. In fig 3 however we observe that the number of outliers is much smaller than the rest of the data points and they might as well have simply occured. Overall I do not think they are enough to throw off the overall model.

Pearson correlation test			
Pair of variables	*Correlation Coefficient*	*p-value*	*Correlation*
cohFac vs blocknum	0.000584	0.8886	No
cohFac vs isDots	0.00486	0.2438	No
cohFac vs isLeft	-0.00207	0.6197	No
cohFac vs RT	-0.133	2.2e-16	**Yes**
blocknum vs isDots	0.00774	0.06309	No
blocknum vs isLeft	-0.00314	0.4505	No
blocknum vs RT	-0.0265	1.892e-10	**Yes**
isDots vs isLeft	0.00259	0.534	No
isDots vs RT	0.5377	2.2e-16	**Yes**
isLeft vs RT	0.005995	0.1502	No

Fig.4 - Correlation table.

From the table in fig. 4 the Pearson correlation test results can be analysed. Correlation occurs between the variables cohFac and RT, blocknum and RT, as well as between isDots and RT. The correlation coefficient between cohFac (number of dots moving in the same direction) and the RT (reaction time) indicates a weak negative association between both variables - if less dots move in the same direction the reaction time decreases, as expected. There is also a negative weak correlation between blocknum and RT - being tired lowers the reaction time slightly. There is a strong positive correlation between isDots and RT which seems to indicate a significant relation between the type of experiment (dots vs control experiment) and the reaction time of the participants - this is also expected, since the control experiment makes it very simple and easy (faster) for participants to select an answer.

3. Model Comparison

According to stepAIC, the final model does not require the variables blocknum and RT. In the first analyses it was hypothesised that the variable blocknum was the least relevant but the variable RT was not expected to be neglected:

```
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
ER ~ RT + isDots + isLeft + cohFac + blocknum
```

```
Final Model:
ER ~ isDots + isLeft + cohFac


        Step Df  Deviance Resid. Df Resid. Dev      AIC
1                              2542   1370.570 1382.570
2 - blocknum  1 0.1503048      2543   1370.721 1380.721
3       - RT  1 1.7356969      2544   1372.456 1380.456
```

b)

isDots: 1 indicates dots trials and 0 indicates control condition.

4 Extending this to algorithms more generally

Yes, an AIC can be used since it would estimate the relative quality of the human detecting models. One must be careful though, for the AIC can process unwanted random attributes - attributes with no actual correlation to the data.

Overfitting can still occur when reusing the same data (used for cross-validation) to optimize analysis parameters. After the cross-validation was deployed, the analysis utilized to classify the data can't be altered to obtain higher performance on that same data. Not doing this (altering the data) can result in overfitting - when a parameter or a classifier goes through a modification.

The lock box method refers to a set of data that is held-out from the optimisation process. This set is used for verification and isn't consulted until the methods have been completely determined. The method doesn't prevent overfitting but rather allows for the testing and verification if overfitting did occur. Cross-validation involves taking a sample of the data multiple times, creating complementary subsets. The researcher then analyses each subset: the training set, and validates the analysis on the other subset, the validation set or testing set. To reduce variability, cross-validation is done various times using different samples, and the validation results are averaged over the rounds to give an estimate of the model's predictive quality.

The blind analysis method - where the input data is altered reversibly in order to avoid that the optimisation process does not tune itself to noise in the data. The pre-registration approach provides proof that the analysis hyperparameters have been finalised before attempting the analyses via other methods. All of the data can be used in the end although hyperparameter optimisation is not permitted.