# Machine Learning CBC - Assignment 5: Case Based Reasoning

Chris Carter (cc1808), Mark Law (ml1909), Michael Thorpe (mt2309) and Fraser Waters (fjw08)
(Group **32+eps**)

For all analysis we have chosen to use the results for the 6 output neural network, rather than the 6 single output neural networks because it was more accurate.
In a particular fold a value of NaN means that there were no occurrences of this emotion. You can find the f1 measure per fold data in the separate spreadsheet, “ML5 Stats.xls”.

**Clean Data**

| | 1 | 2 | 3 | 4 | 5 | 6 | Avg |
|---|---|---|---|---|---|---|---|
| Decision Trees | 0.7067 | 0.9524 | 0.8800 | 1.0000 | 0.9238 | 0.9630 | 0.9043 |
| NNs | 0.8651 | 0.8429 | 0.9333 | 0.9133 | 0.7333 | 0.9778 | 0.8776 |
| CBRs | 0.9111 | 0.9857 | 1.0000 | 1.0000 | 0.9714 | 1.0000 | 0.9780 |

As shown in the results above our Case Based Reasoning algorithm performed best on the clean data set. However this does not show that this is the best algorithm in general as this is one very small data set and the parameters that we have chosen (ie the metric and the number of nearest neighbours to consider) were selected to give the best results for this data. In order to determine whether this algorithm truly is better than the other two we must do further tests on a new data set using these same parameters.

**Noisy Data**

| | 1 | 2 | 3 | 4 | 5 | 6 | Avg |
|---|---|---|---|---|---|---|---|
| Decision Trees | 0.7833 | 1.0000 | 0.9407 | 0.9800 | 0.8056 | 0.9249 | 0.8774 |
| NNs | 0.8267 | 1.0000 | 0.8125 | 0.9101 | 0.9000 | 0.9026 | 0.8623 |
| CBRs | 0.8667 | 1.0000 | 0.9407 | 1.0000 | 0.9444 | 0.9506 | 0.9380 |

The best learning algorithm, based on the above results, is the Case Based Reasoning system. This is also the most consistent algorithm as is it uses the least ‘randomness’. Given the same training and test data, there will be very little change in the average f1 measure when run multiple times. However for the decision trees the average varies significantly.

## Statistical Analysis

We ran the ANOVA, Multiple Comparison and t-tests on both the clean and noisy data. The results can be found in the separate statistics sheet.

We first considered the ANOVA test. Mainly we looked at the p-value. This is the probability, given the null hypothesis, that F is as extreme or more extreme than the value returned for F. The lower the p-value the more evidence that the algorithms give different results.
If a case presented with a significantly low p-value we considered the results of the t-tests and the multiple comparison tests.

NaN in the results for t-tests indicate the two algorithms were identically distributed for this data. Similarly NaN in the results for ANOVA tests indicate that all 3 algorithms had identical distributions for this data.

**Clean Data**

For the clean data, four out of the six emotions give F-statistics which are greater than 1 and 3 of the emotions give p-values lower than 5%. This shows that we needed to consider further statistical analysis on these emotions.

Upon further analysis the t-tests showed, at the 10% significance level, that four of the emotions showed a difference in the distributions of the Neural Networks and the CBRs. Two emotions also showed that the Decision Trees were differently distributed to the Neural Networks. And one emotion showed a difference between CBRs and Decision Trees.

When considering the results of these emotions under the multiple comparison test two emotions show that Decision Trees are better than the Neural Networks and four emotions show that CBRs are better than the Neural Networks.

The tests overall show that both the CBRs and the Decision Trees perform better than the Neural Networks. There is limited evidence to show any difference between the CBR’s and the Decision Trees. However if we were to up the significance level to 15% the data suggests that CBR’s perform slightly better.

**Noisy Data**

For the noisy data the ANOVA tests indicated that emotions 3 and 4 required further investigation. However the t-tests only showed that CBRs and Neural Networks might have different distributions and the multiple comparison tests seem to show that at the 10% significance level there is not enough evidence to show that the algorithms perform differently.

## Adding New Emotion Classes

Adding new classes of emotion to the three algorithms would require some small changes: this (very short) bash command would be perfectly functional in replacing all instances of ‘6’ with ‘7’ sed -i 's/6/7/g' *.m, as all the instances of 6, would simply need to be replaced with 7. However, given that we have optimised the input parameters for the neural networks to produce the bests results for 6 inputs, it seems likely that we would have to make several alterations, in order to get the same performance. For the CBRs, it is possible that the k-nearest-neighbour parameter would also need changing for the best performance.

The decision trees algorithm would require the least tweaking as there are no parameters required, however the algorithm is currently hard-coded to create and expect 6 decision trees. This would obviously change with more classification possibilities as there is one tree per class.

## Analysis of results

The ANOVA test assumes that the samples are normally distributed, with equal variance, and that each sample is independent. The t-test also assumes that the samples have equal variances. To avoid this assumption we could have used the Welchs t-test.

Overall the Case Based Reasoning algorithm seemed to perform best especially on both the clean and noisy data. It was also the most consistent algorithm (because it doesn’t depend as strongly on random choices as the others do). The NNs did not perform as well, and this may be due to the fact that they were calibrated to suit the clean data set, and therefore the attributes of the network may be suboptimal for the noisy data set. The decision trees’ performance varies greatly for each run of the data, however on average it performs consistently worse than the CBRs.