**Comparison of algorithms II:**

I have already compared all the classification algorithms against each other in the previous document.

The rest of the supervised learning algorithms are: regression, neural nets, and support vectors. I used the same dataset for them all. With regression I had an adjusted R-squared value of 0.745. With neural nets I had a correlation of 0.69. With support vectors 67% of my data was predicted correctly. I don’t know how to compare adjusted R-squared value to correlation and percent data predicted correctly.

Neural nets and support vector have similar performance. However, with neural nets, with different seed values the correlation went as low as 0.2 so I think support vectors are more reliable than neural nets in this case.

Regression is good in that it is not black box like support vector and neural nets. But it isn’t as precise as them with bigger datasets either. Also with regression, there is a lot of work in figuring out the significant predictors and making decisions about if some predictor parameters should be squared etc. so the skill and judgment of the statistician is important whereas with neural nets and support vectors that’s not an issue.

However with the automobile insurance dataset, it turns out that clustering, which is an unsupervised learning algorithm, may have been the most useful since the most significant predictors as found in regression, was “make” which is a categorical predictor and regression, neural nets and support vectors are more suited to numerical predictors. Using clustering, the data can be divided into different clusters for make, and different models can then be generated to predict the target variable value for different clusters.

Finally, I used a completely different dataset for association rules, the second supervised learning method. I got rules with good lift values but felt very blind working on it as I didn’t know the significance of the rules and would need subject matter experts.

Both the unsupervised learning algorithms can be used for very different purposes from the supervised learning algorithms but it is hard to quantitatively judge their performance and it seems they need more qualitative attention.