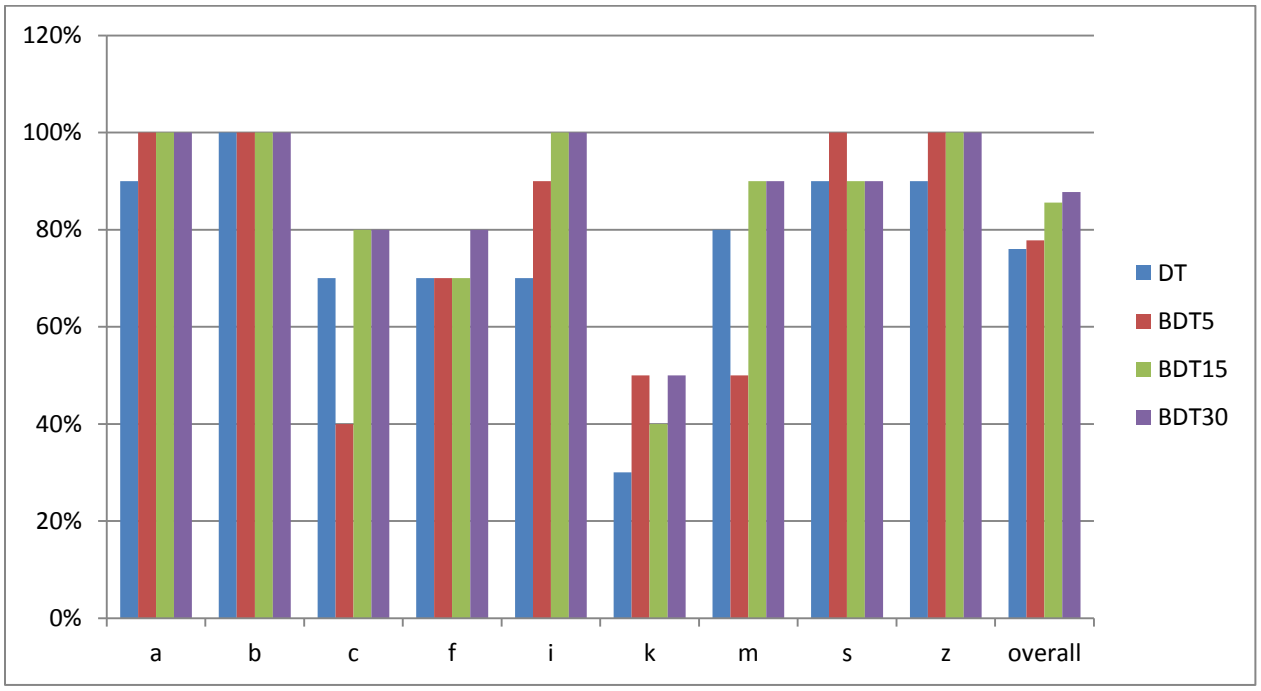
John Nickle

Decision Trees and Baggers

Decision Trees turned out to be somewhat successful, but the real takeaway from this project was the power of Baggers. After implementing bagging on my decision trees, writing recognizers, and SOMs, they all had a great increase in accuracy! This was especially true for the SOM as with 30 bags it had a 100% success! Bagging seems to be the answer to conforming your data into reliable results.

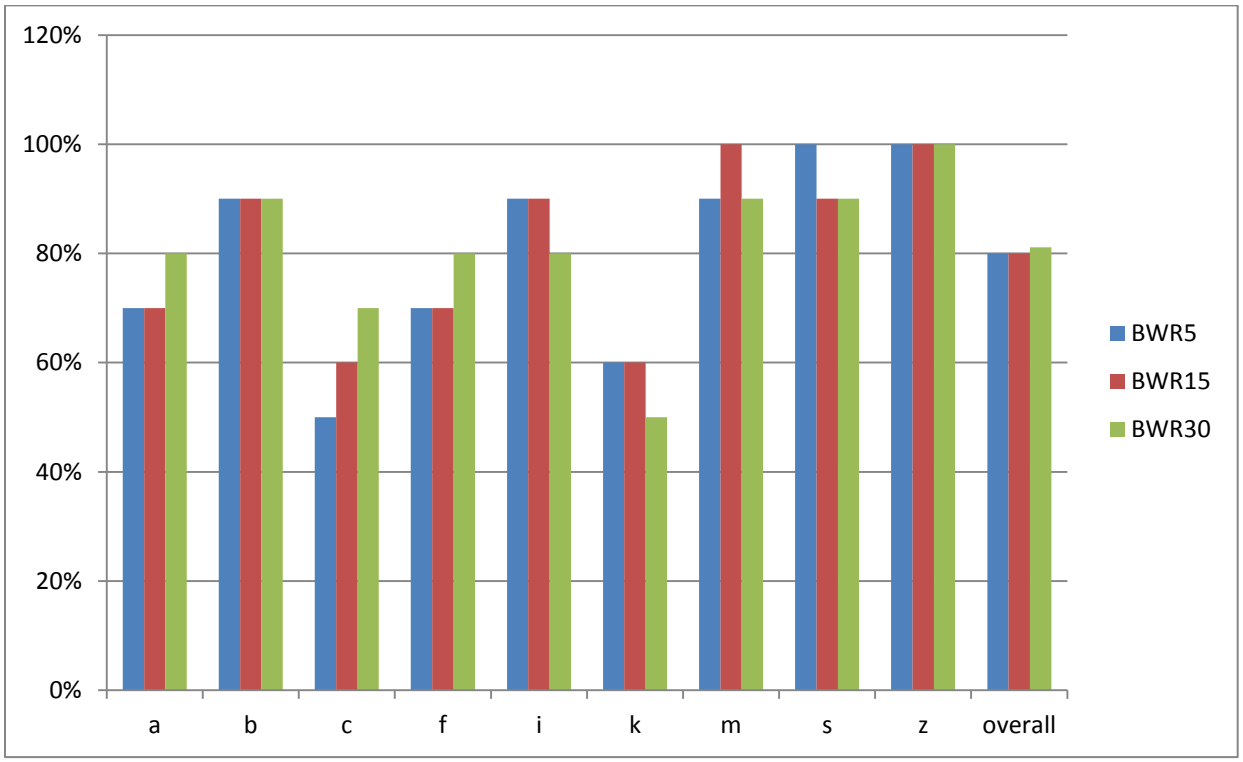
The data from my Decision Tree follows:



As you can see, the decision tree did decent by itself, but bagging increased the accuracy to nearly 90% overall! In examining the data, I wondered why k was never very successful. Looking at the drawings, about 4 of them are poorly drawn. This certainly could hurt the training and therefore the ability to classify better drawn ks. With redrawn data, I'm sure the results would be even better.

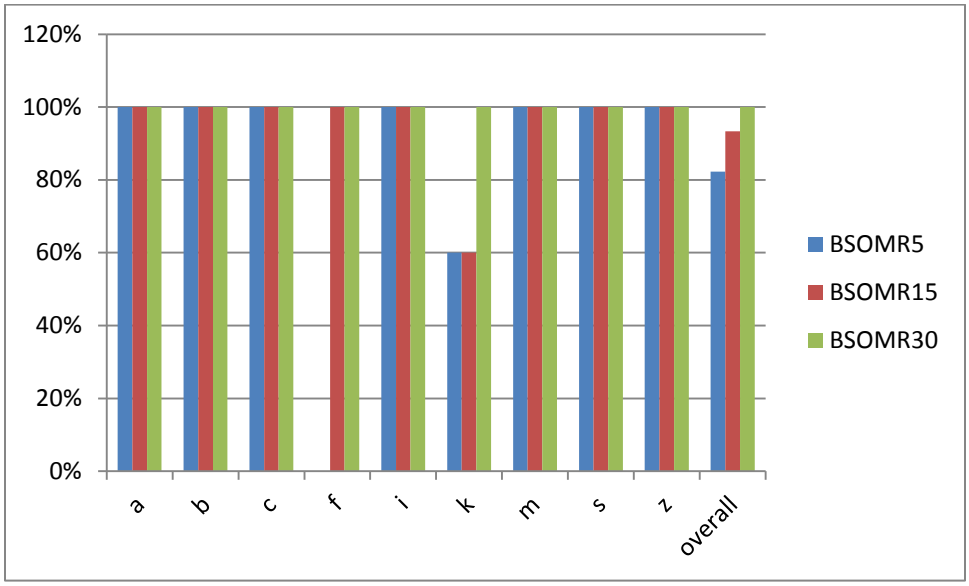
As stated initially, Bagging is the real achievement of this project. Bagging increased the results of every learner, but only SOMs in a significant amount. The bagged SOM even reached 100% accuracy with 30 bags!

Here is the data for BaggedWritingRecognition:



Examining the data, there is not much of an improvement with bagged WRs. This could very well be due to the way that each picture is encoded and classified; randomizing the data doesn't really have an impact on that. Because of this, bagging should not really be considered when dealing with MLPs.

The results from my SOMs were simply amazing:



At 30 bags, it achieved a100% accuracy; I even checked it by manually drawing symbols for it to classify! Upon closer examination, we can see that the SOM receives the most benefit from bagging. SOMs drastically alter when given different data sets and even different orderings of data sets. It is easy to see why SOMs would be most benefitted from bagging.

Overall, SOMs outperformed the other learners by a significant amount. This was due to bagging and the seemingly logarithmic affect it has on the success rate by changing up the data samples. DTs did average, as did the WRs. As a result, I think bagging is only really helpful for unsupervised learners such as SOMS, where every little detail can change the outcome of training. Looking at both DTs and WRs, bagging did not really affect the results in a meaningful amount; the tradeoff of a 30x longer algorithm to increase results by 1% is simply not worth it.