

Final Project – Personal Reflection
BMEN 415
Jeff Roszell
30043087

In our final project we have decided to work on three datasets. Our first was the birth weight prediction dataset. As this is a numerical output which is singular, we chose to use this as our regression-based solution. The second was the Alzheimer's database, in which there were three classifications or targets. Based on this we decided that we should use the classification models in order to come up with predictions. And finally, our challenge case, was recognition of certain ophthalmologic conditions based on an image dataset.

Methods

Individually for the birthweight I decided on three regression models. MLR was the simplest that I used, then I moved onto Lasso, and then finally a mix of ridge regression with Bayesian network flavour.

For the classification problem I decided to use three models. kNN was my first and least involved, then the SVM algorithm, and then finally the Multi-Layer Perceptron neural network.

Predictions

For the regression I hypothesized that as you progressively make the models more involved, they would also become more accurate. Therefore, I thought that MLR would be the least likely to be accurate.

Similarly, I thought the same with the classification problems. However, I knew that SVM was very well suited to the problem, so I was hoping that it would be able to beat MLP as it is a black box and no one ones how that works.

Outcomes

In the regression models the results were:

MLR 7.17835745600906E+23

Lasso 5.50893154887483

Bayesian Ridge 0.132693791884719

In the classification models the results were:

kNN 0.788732394366197

SVM 0.873239436619718

MLP 0.887323943661971

The regression results are a measure of MSE and the classification results are in accuracy.

Reflection

From these numbers I can see that my initial thoughts on which of the algorithms would be most suitable were correct. I was sad to see the MLP outperform the SVM for the classification model. I believe that with further pre-processing of the data I could get the accuracy up for the models.

Reflections on the course

This course being the first in its history, was good. I feel that there were some definite growing pains in the beginning, some soul-searching even. Eventually I personally was able to grasp how I needed to learn the topics, which included the lectures, YouTube videos, and some textbook reading to follow up if required.

The topics presented were almost entirely machine learning driven. If that is the intent of the class, then great. If not, then this is just a heads up. The math seemed daunting however I get where it was important to approach these topics in order to further understand what we actually building with the machine learning systems.

Contribution to the group portion

Initially helped with the “bones” of the CNN up to a point which others could try out their own parameters and adjustments. Additionally setup the gray scaling pre-processing effort in order to reduce the dimensionality of our dataset. During the collection of the overall data, I was able to help make the graphs depicting the results of MSE for the combined systems. Helped others with errors on a consistent basis. At the end, because of our low accuracy and level of prediction, I spent many hours writing and rewriting the CNN from scratch in order to try different approaches and classifiers in order to help come up with the final product.