Maria Eugenia Romero

April 5, 2021

Dr. Eick – 4368

Task 4

Using SVM and NN Tools

Data:

Heart failure clinical records from the UCI Machine Learning Repository

Findings:

Classifier	Params	Mean Accuracy 10-Fold CV (STDDEV)
SVM – Linear	N/A	.800 (0.085)
SVM – RBF	C = .78	.679 (.066)
MLP – ReLU	Learning Rate = adaptive, random_state=100, max_iter = 5000	.676 (0.067)
MLP - Logistic	Learning Rate=adaptive, random_state=100, max_iter= 5000	.679 (0.066)

Report:

The best performing classifier for this data set was a Linear Support Vector Machine. This model had a mean accuracy using K-fold cross validation of .800 or 80%. I believe this model outperformed others because an SVM simply categorizes data points, and the model decides what a good classification for each point would be for the given data. Rather than an MLP, which would be more used in the case if we needed to ***learn*** about what causes deaths. However, because this data is tabular and is observed data, an OLS Linear Regression can also be used to make sense of the observations in the medical findings—therefore making SVM Linear the optimal tool for beginning analysis with the tabular set.

The first most important finding of this task is that complex does not always mean better classifier/neural network. Because, in the case of the MLP, all else equals, logistic is a bit more functional than ReLU in terms of mean accuracy for K-fold cross validator. This is because logistic, is best used in the final layer for binary classification problems. And that is the case for our data that were working with where the Y is a dummy variable, 1 – death, 0 – no death.

The second most important finding was that models and data need to match. Not all data needs to be learned, some data works best classified. That is what I believe the case is here, because our 10-fold cross validation held strongest with SVM linear, that SVM linear works best with smaller data. This can also be an explanation as to why MLP had such low accuracies—because of the need for a bigger data set to successfully finetune an MLP NN. There needs to be more than just 299 data points to full train and test and learn and increase the recall for a neural network.