Ramatu Yusuf, Emily Klassen, Evan Duffy, Melody Majoni

March 3rd, 2020

ADMN 5016

Professor Babaei

An Empirical Comparison between Deep Learning, Boosting, and SVM Machine Learning Models

We will be comparing Deep Learning, Boosting, and Traditional Machine Learning Models (SVM and Tree methods) across 3 different datasets. The datasets were strategically chosen to present different kinds of classification – two binary classifications, and one multiclass classification problems. The datasets are of very different size, ranging from 210 rows in the Wheat dataset to 30,000 rows in the Credit Card dataset. The binary classification datasets are unbalanced, with more cases in the negative class than the positive class. The Wheat dataset is perfectly balanced.

The cancer dataset was the same one that was used in class for our first individual assignments. The dataset presents columns with different analytical measures of various tumors and classifies them into benign or malignant tumors. Although favoured toward the negative class, this dataset presents a relatively balanced spread of data, with 357 benign cases and 212 malignant cases. We chose this dataset because each member has previously analyzed it with classic Machine Learning techniques, and so it would be interesting to revisit the dataset armed with new techniques. Thus, our will also have the contextual familiarity having analyzed it with basic machine learning models.

The wheat dataset is smaller, with only 210 rows of data, divided into kernel data used to determine the type of wheat the kernel belongs to. Like the iris classification dataset, the dependent variable is classification of 3 different types of wheat. With each variety comprising 33% of the rows, this is an evenly balanced dataset that would provide a different outcome than the comparatively unbalanced cancer dataset.

Our third choice is the Taiwan Credit Card dataset, which attempts to predict credit default status under certain criteria such as amount of credit given, marital status, bill amounts across time, etc. The dataset is heavily skewed to the negative class – that is, non-defaults— and may therefore be an interesting case to run oversampling looking to better determine default cases. Compared to our other two datasets, this is much larger, at 30,000 rows.

The boosting algorithms we have chosen are XG Boosting and Random Forest, which are both Tree-based; our Deep Learning model is Neural Network, and our classic Machine Learning algorithms are SVM and Decision trees. By comparing algorithms of different evolution across different types of datasets, we hope to test the ‘no free lunch’ theorem, which asserts that there is no one specific algorithm that will be the best solution across all datasets. Each model comes with its strengths and limitations; and, depending on one’s judging metric, each algorithm will have a different outcome.

We will know when the project is successful after we have made a fair comparison across all our datasets with all our models. We will use a ROC-AUC curve as our evaluation metric to gauge performance. We will also look at overall accuracy and F1-scores, to see how each algorithm performs with span of small and large datasets with both the balanced and imbalanced datasets.

Given the recent excitement surrounding Boosting methods such as XGBoost, we expect them to perform better than more traditional methods such as classic Decision Tree.  Or, perhaps we’ll see that we should be cautious when buying into hype.