NEURAL NETWORKS

Michael Brosnan

CIS 435 PRACTICAL DATA SCIENCE USING MACHINE LEARNING

PROF KAKADE

Business Problem

The business problem brought to us on the data team is to build an algorithm to determine if a particular patient's breast cancer is benign or malignant. We were able to pull publicly sourced data to put together a dataset of 569 cases of breast cancer with the contributing variables and whether the cancer was malignant or benign. This data set will be the basis of the model we train and test. While this is similar to the business problem we encountered in a previous assignment, this time we will be using neural network and deep learning machine learning to attempt to produce better results than the clustering models used previously.

To answer the above business problem, we will be using the CRISP-DM method, or the Cross Industry Standard Process for Data Mining, which is an open standard process model used for data mining and machine learning algorithms. The steps of the process include business understanding, data understanding, data preparation, modeling, evaluation, and deployment. We have already begun this process with the business understanding, or understanding the problem which the business is trying to solve, which in this case is to use a neural network-based model to analyze certain inputs and determine if a patients cancer is benign or malignant. We will then begin understanding the data and making sure we can answer the business problem with the data we have, and adjusting if necessary. We will then prepare the data, build out our model, evaluate the model against our business understanding, then deploy the model once it has been properly vetted and tested. Using these steps will help us deeply understand the problem and ensure that the result we produce is in line with what the business wants and is effective in solving their problems.

Machine Learning Applications

Healthcare is an interesting area when discussing machine learning because there is often hesitancy to implement machine learning applications because of the lack of understanding by end user on how the machine learning algorithms actually work. Plus, the fact that there is often times a human life at stake make using machine learning a controversial topic. While there are some organizations that are hesitant, many others are embracing the power of machine learning and implementing different models to help enhance their care and provide providers with more information and intelligence that they may not have otherwise. We will discuss neural networks in greater detail later, but because neural networks are more of a black box algorithm compared to regression or clustering algorithms, there is more concern about implementing these models. However, neural networks have been a positive influence on a number of healthcare organizations; and the below examples offer a brief overview of the different areas neural networks can improve an organization. These organizations that embrace machine learning and neural networks have a competitive advantage against other area competitors, who may not be able to match organizations using machine learning and neural networks to produce better results.

Neural networks are becoming more widely used in the healthcare environment, and one of those uses is detecting a diabetic eye disease in patients. Diabetic retinopathy is “a diabetes complication that affects eyes. It's caused by damage to the blood vessels of the light-sensitive tissue at the back of the eye (retina) (*Diabetic retinopathy* 2021).” It is quickly becoming a massive issue in the healthcare community and is the fastest growing cause of blindness in the world, with 415 diabetic patients at risk of developing the disease across the world. Many of these patients are in countries where specialist medical care such as ophthalmologists are not readily available, so often times this condition goes unnoticed, which causes the disease to worsen to the point of blindness (Peng & Gulshan , 2016). The main way of detection for this condition is a specialist eye exam, and the images produced from that exam require a specialist to identify diseased images. A group of doctors decided that they would attempt to use machine learning, specifically neural networks to help doctors in areas with a lack of specialist support identify diabetic retinopathy quickly and efficiently so patients could receive care as soon as possible. With the help of doctors from India, a dataset of 128,000 was created and fed to a neural network machine learning algorithm to allow the algorithm to learn and identify those diseased images. Overall, the F-score, a score to determine the accuracy of the model, of the model was .95, beat the F-score of eight ophthalmologists who were used as the baseline (Peng & Gulshan , 2016).

Another interesting example of neural network used in healthcare centers around mental health outcomes. This is particularly interesting because mental health is a difficult area to quantify, with the brain being an area that we are still learning about every day. Recent advances in machine learning and neural networks have been shown to provide better performance than previous versions; and they are now able to help psychiatrists and psychologists make better decisions for patient using their data- specifically medical data, behavior health data, and even social media data (Su et al., 2020). So far, because of a lack of real data, most neural network and machine learning applications so far in the mental health arena have been focused on the classification of psychiatric disorders, such as dementia. There are also more studies on the diagnosis of schizophrenia, autism, and depression, although some of the potential indicators of those are harder to find (Durstewitz et al., 2019). More research is being done around depression and other mental illness in teens and young adults; with interesting ideas of how to find data also being discussed. Rather than focus on some of the data inputs that can be somewhat unreliable like self-reported data, some models look to used data from social media companies such as Reddit and Twitter to identify mental health indicators. There is also interest in using data from mobile devices such as wearable devices and mobile phones to identify mental illness (Durstewitz et al., 2019). These different data sources could add to the potential of neural network-based algorithms, and they may be able to identify early on if patients are descending into mental illness and take steps to pull them out.

The final example of a neural network used in a healthcare setting is predicting the length of stay (LOS) of a patient in the Emergency Department. A group used neural networks from over 16,000 patient encounters to develop a model that could predict how long a patient would be in the emergency department before discharge or admission. This is important because the emergency department is often used a primary care facility for patient who do not have insurance as well as the homeless and older individuals. With limited inpatient bed availability, the emergency department can often become a madhouse with all the people trying to get a bed. Understanding the amount of time a person is likely to be in their bed allows bed planning and the hospital time to strategize on how they will deal with the waiting individuals. This specific model did not do a particularly good job of predicting the time a patient would stay, with an average of within 7.5 hours on the testing set (Wrenn et al., 2005). While this specific model did not perform well enough to be implemented yet, there is promise in the development being done, and the providers expressed confidence that underlying patterns were recognized and could be built upon.

Neural networks can be a power tool in healthcare and can improve results when used in the proper settings. One thing that stood out when looking at actual examples compared to previous model like regression or clustering is that neural networks seem to be more of a work in progress. Many of the examples I looked at when researching this topic were that some of the results were not perfect well-formed models, but that there was excitement and confidence that neural networks would be improved upon and will be implemented soon as more and more work is done with these models.

Machine Learning Algorithms

To solve the business problem at hand, we want to build a neural network machine learning algorithm that can take in inputs and produce a prediction based on that set of inputs. A neural network, sometimes also called an artificial neural network, is a machine learning algorithm that is designed to mimic the way the human brain works and makes decisions. In the mathematical sense, an artificial neuron is “a nonlinear transformation unit, which takes the weighted summation of all inputs and feeds the result to an activation function, such as sigmoid, rectifier (i.e., rectified linear unit [ReLU]), or hyperbolic tangent (Su et al., 2020).” These can also be referred to as nodes. Neural networks are made up of layers, with the number of layers being variable. It begins with the input layer, then any number of hidden layers which are the deep neural network layers, and then the output layer (IBM Cloud Education, 2022). Each of these layers is consisted of nodes, which are connected to one another and have different weights associated with them, as well as a threshold. When the output of a specific node is higher than the designated threshold, the node is activated, and the data is passed along to the next node in the process. A neural network depends on training, which is done through running large amounts of data through the model so it can learn from the data and be able to make accurate predictions. There are subsets of neural networks such as Recurrent neural networks, convolutional neural networks, and autoencoder that can be used for more specialized use cases.

Recurrent neural networks (RNNs)were created and designed to recognize and analyze speech, video, natural language, and other sequential data. The RRN takes in input elements one at a time to a recurrent node, and each node that receives an input also receives a date and time stamp associated with that element (Su et al., 2020). The part that brings the different neurons together is the recurrence link and that allows the RNN to capture dependencies of other elements as well as develop the sentence structure. Other varieties of RNNs have been introduced in the past few years, including long short-term memory and gated recurrent unit, with the difference in those models being how the input and output are connected in order to make the output make sense. These models are very useful when attempting to analyze clinical notes and social media posts to help identify possible signs to certain mental health conditions (Su et al., 2020).

Convolutional neural networks (CNNs) are another interesting deep machine learning model that is and can continue to be very important to the medical field. This type of neural network is used for image classification, as they take pixels from images as inputs to the model and map them to the corresponding target. The convolutional neural network is usually split into three types of layers, convolution–activation layer, a pooling layer, and a fully connected layer (Su et al., 2020). The activation layers take the inputs and convolves them into two-dimensional convolution filters, which are taken to the pooling layer, which reduces the size of the inputs. There is then the hidden layer that connects the connects each neuron to all the other neurons in the previous layer (Su et al., 2020). In the previous section we went over an example of using neural networks to identify disease, this is the type of analysis that can be done using CNNs.

The final neural network example we will discuss is autoencoder, which is used for feature learning for unlabeled data. The autoencoder typically consists of two parts, the encoder that does much of the feature learning and identification, and the decoder which does the opposite, and rebuilds the data back to what it started at from what was taken from the encoder. Autoencoder is especially useful when you have raw data and would like to perform unsupervised learning on that data (Su et al., 2020).

Neural networks can be extremely helpful and powerful tools. They are flexible and can solve both regression and classification problems. They are also good for nonlinear data with large amounts of inputs, such as voice and picture data. These models work quickly and work best with more data points. Unfortunately, they also have some downsides, one of the biggest being that they are black box algorithms, so it can be difficult to determine the effect each input on the output. The other downside is that neural networks very much depend on the training data inputted, which can lead to overfitting because the node learns from that data and thus may be more tuned to that training data (Ciaburro & Venkateswaran, 2017). Similar to the black box issue is the issue of explain ability of the neural network model. Models we looked at previously, like regression or the supervised clustering, are more easily explained as they are more of mathematical models and can be broken down and understood. This is unlike the neural network models, which are mathematical in nature as well, but in between the input and the output there is a lot more of an unknown aspect, as they are not able to be investigated and can be difficult to explain. This is the trade off with power and performance, as the easily explainable models tend to not perform as well as the less explainable models like neural networks.

Data Preprocessing Discussion

The data preparation for this business problem was very similar to previous business problems we have encountered. We first imported the necessary libraries into the Jupyter Notebook, which included pandas, numpy, matplot and sklearn. We then read in the data set and assigned the X and Y values, with the Y value being the dependent variable of diagnosis. Once we did that we learned more about the dataset, including the count, mean, standard deviations and if there were any null values. From there we identified the type of data that each column was; and at that point we were confident of the preparation that had been done and were ready to build out our model. We also split out the data into training and testing sets, the training would be used to feed through the model so it could learn from the data, then we would test the model using the test data and evaluate its performance using that set.

Explaining Metrics

The metrics used to determine the success of this model were very similar to that of the clustering supervised learning that we discussed in a previous assignment. The first items we looked at were the accuracy of the model on the training and testing set. This basically shows us how well the model predicted the correct outcome of if the patient's cancer was benign or malignant. We concentrate more on the test set, because hat contains data that the model was not trained on.

The second metrics used will be the confusion matrix. The confusion matrix separates the output results from the model into four categories, true positive, true negative, false positive, and false negative. The first two are result where the model does well, in a true positive, the model predicts positive and the patient is positive. The opposite of that is true for true negative, the model predicts negative and the patient was actually negative. These are both cases where what the model predicted was the actual outcome, so the model was correct. A false positive is a situation where the model predicts that the result is a positive, and the result is negative. Again, the opposite is true for a false negative, the model predicts a negative, but the result is actually positive. These are times when the model is wrong in its prediction. We will go into more detail on how this model performed in this aspect and how that effects people in the next section.

We will also look at precision, recall, and the F-1 score. Precision looks at the number of correct positive predictions made by the model and divides them by the total number of positives predicted by the model. Recall is the number of correct positive predictions made divided by the total number of positives in the dataset. The F-1 score is then just the mean of the recall and the precision scores.

The final metric we will look at is the area under the curve (AUC), which is the amount of area beneath the curve in a Receiver Operating Characteristic (ROC) graph. This graph shows a line for a 50/50 guess as the result of a dataset, then determines how the model performs in comparison. A strong AUC is closer to one and indicates a high performing model.

Interpreting Results

While interpreting the results, we will compare the neural network model created for this business problem to the clustering algorithms used for the same business problem previously. This will give us a good idea of how the model performed against other algorithms with the same data set being used. We will start with accuracy, precision, recall, and F1 score, then we will dive deeper into the confusion matrix for the neural network model compared to the other models.

	Accuracy on test	Precision (0/1)	Recall (0/1)	F1 (0/1)	AUC
Logistic Regression	.9231	.94/.98	.99/.89	.96/.93	.988
Decision Tree	.9301	.93/.92	.95/.89	.94/.91	.942
Naïve Bayes	.9301	.91/.96	.98/.85	.95/.90	.973
Neural Networks	.958	.97/.95	.97/.95	.97/.95	.989

The above table shows the results of several metrics from the four models that were created for the breast cancer data set. The accuracy on the test set metrics shows us how well the model did when evaluating the separated-out test values. All the models were above 90% accurate, but the three supervised models were in the low nineties, none higher than 93.01%. The neural network model performed much better at 95.8%, which is more than two percentage points higher than the other models, showing that this model is more accurate, and the difference is substantial given the nature of the subject it is predicting. The neural network model also performs better overall in precision, recall, and F1, with slight exceptions in the recall and precision for part of other models, but makes up for it in the F1, which is the harmonic mean of the two others. The final and possibly most important metrics is the AUC number, which shows how well the model performs. The neural network model again performed the best, beating the decision tree by over four percentage points, the naïve bayes by one, and the logistic regression by .1%. This shows that the neural network model is the highest performing model, and it performed very well against the test set provided.

The final important data points to look at is the confusion matrix for the model. The accuracy metric we discussed above is derived from the performance of the model in terms of the confusion matrix, which shows the number of true positives, true negatives, false positives, and false negatives that the model predicted.

Neural Network	Positive	Negative
Positive	85	3
Negative	3	52

The results of the confusion matrix for the neural network model are above. 85 indicates the number of true positives in the data, and 52 indicates the number of true negatives. This means that there were 85 times the model correctly made a positive prediction, and 52 times it predicted negative, the result was negative. This are the instances where the model predicted correctly, and that is the vast majority of the cases, over 95% according to the accuracy test metric. There were only six times in the test set where the model predicted incorrectly, and that is indicated in the lower left and upper right boxes. There were three times when the model predicted a positive result and the result was negative, and there were three times when the model predicted a negative result and the result was positive.

Overall, the model performed extremely well, with only six wrong results in the entire test set, and the 95.8% accuracy is extremely impressive. If this was an industry such as finance or advertising, this would be an outstanding result, and there would be no hesitancy in implementing the model and taking insights from it. The business problem at hand is in the field of healthcare though, specifically related to identifying benign and malignant cancers in patients. While the 95.8% is impressive, there are still 4.2% of case where the model predicted a wrong outcome. That means that when an individual gets the results of their test back, whether the answer is malignant or benign, there is a 4.2% chance that the result is incorrect. If patients get a result of benign, mean that it is not as much of a threat at a malignant tumor, there is a 2.1% chance that that reading is incorrect, and they are actually in danger and do not know it. The opposite is also true, there is a 2.1% chance that an individual gets a reading of malignant when it is actually benign, and they will have to undergo treatment when it is not completely necessary. This is the interesting conundrum in the healthcare industry, this model performs extremely well, better than many other models in production in other industries, but the threshold for using a model in healthcare is so high because human lives are in the balance. In other industries, a 95% accuracy in a model is a no brainer, but in healthcare, it might not make the standard necessary to use.

Recommended Steps

The recommended steps for this model follow closely to what was said in the previous section and to what was said when reviewing supervised learning models: the standard and stakes in healthcare are so high that if you are going to implement a machine learning model it must be unbelievably accurate, with an over 99% accuracy rate. The model we create, while very high-performing, still leaves questions for about 4% of the population. My recommended next steps would be to gather more data and continue to train the model so that it can continue to learn from the inputs and become even more accurate. Once it has a suitable accuracy rate, whether that be over 98% or 99%, implement the model on a trial basis, with the results reviewed by physicians, and compare the physician results to the results of the model. When the data team and the clinical team are comfortable with the results, you can implement the model on a full-time basis, but the model must continue to be fed input data and reviewed so that the accuracy levels stay high.

References

By: IBM Cloud Education. (n.d.). What are neural networks? IBM. Retrieved February 20, 2022, from https://www.ibm.com/cloud/learn/neural-networks

Ciaburro, G., & Venkateswaran, B. (2017, September). Neural Networks with R. Retrieved February 20, 2022, from https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781788397872/1/ch01lvl1sec27/pros-and-cons-of-neural-networks

Durstewitz, D., Koppe, G., & Meyer-Lindenberg, A. (2019, February 15). Deep Neural Networks in psychiatry. Nature News. Retrieved February 20, 2022, from https://www.nature.com/articles/s41380-019-0365-9

Mayo Foundation for Medical Education and Research. (2021, June 24). Diabetic retinopathy. Mayo Clinic. Retrieved February 20, 2022, from https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611#:~:text=Diabetic%20retinopathy%20(die%2Duh%2D,or%20only%20mild%20vision%20problems

Peng, L., & Gulshan , V. (2016, November 29). Deep learning for detection of diabetic eye disease. Google AI Blog. Retrieved February 20, 2022, from https://ai.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html

Su, C., Xu, Z., Pathak, J., & Wang, F. (2020, April 22). Deep learning in mental health outcome research: A scoping review. Nature News. Retrieved February 20, 2022, from https://www.nature.com/articles/s41398-020-0780-3

Wrenn, J., Jones, I., Lanaghan, K., Congdon, C. B., & Aronsky, D. (2005). Estimating patient's length of stay in the emergency department with an Artificial Neural Network. AMIA ... Annual Symposium proceedings. AMIA Symposium. Retrieved February 20, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560706/