Machine Learning For Cryotherapy

Milesh Shah

Abstract

Many methods in machine learning are now being used for many industries especially in the medical industry where they are being used for tasks like disease diagnosis and predicting the result of a treatment. Multiple machine learning approaches have been used on the classification of Cryotherapy, where accuracies ranged from 59 to 95 percent, sensitivities ranged from 82 to 96 percent, specificities ranged from 44 to 98 percent, AUC (Area Under the Curve) ranged from 58 to 95 percent and F1 Measure ranged from 64 to 94 percent. The best performing model from that study was the Random Forest, achieving an accuracy of 95 percent, an F1 Measure of 94 percent, a sensitivity of 88 percent, AUC of 95 percent and a specificity of 77 percent. In this study, the classification approach that will be used is Logistic Regression. After analyzing this method, this model achieved an accuracy of 86.67 percent when validating the machine learning model on a test data set containing 15 samples and it achieved an accuracy of 73.34, a sensitivity of 69.75 percent, a specificity of 80.07 percent, AUC of 77.22 percent and an F1 Measure of 72.08 percent when validating the machine learning model using 10 fold cross validation. While this approach outperformed some of the previous approaches, this study will discuss the drawbacks of using Logistic Regression for classification on Cryotherapy.

1 Introduction

Today, in the medical industry machine learning has found many applications and one of them includes using machine learning models to predict the potential outcome of a medical treatment. In this study, a cryotherapy data set provided by Khozeimeh F et al [3] in the UCI Archive KDD [1], is used to predict if warts can be healed by the treatment method known as cryotherapy. This data comes from a dermatology clinic of Ghaem Hospital in Mashhad, Iran. The data set includes information about 90 patients. The information of each of these patients that is included in this data set includes the patient's age, sex, number of warts, time elapsed before treatment, area of warts, the type of warts, and whether the treatment was successful in healing the warts. The data set includes information of 42 patients for whom the treatment was not successful in healing the warts and 48 patients for whom the treatment was successful in healing the warts.

In a previous study conducted by Ali Cuvitoglu and Zerrin Isik [4], shows their results where they used multiple machine learning approaches for predicting the result of the cryotherapy treatment. In that study, the methods they explored included Naives Bayes (NB), Random Forest (RF), Support Vector Machines (SVM), k Nearest Neighbor, and Artificial Neural Network. They also explored some Dimension reduction and feature selection methods such as Principal Component Analysis and Linear Discriminant Analysis to see how they performed in comparison to approaches where there was no dimension reduction or feature selection. Another approach that was used was one called ordinal classification and ensemble, which is using Naive Bayes, Random Forest, Support Vector Machines, k Nearest Neighbor, and Artificial Neural Network and using a voting system to decide to which class the sample belonged. These machine learning approaches managed to achieve ac curacies ranging from 59 to 95 percent, sensitivities ranged from 82 to 96 percent, specificities ranged from

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

44 to 98 percent, AUC (Area Under the Curve) ranged from 58 to 95 percent and F1 Measure ranged from 64 to 94 percent.

In this study, Binomial Logistic Regression will be explored as an approach for predicting whether the cryotherapy treatment will be successful in treating the warts and it will be compared to the most optimal and the least optimal approaches that were used in the previous study. The reason for using this method is that in the data set whether the treatment will or will not be successful in treating wart is denoted as a 1 for saying that the treatment will be successful in treating the warts and a 0 for saying that the treatment will not be successful in treating the warts. That makes this a classification task with two classes for which logistic regression is a common approach.

2 Method

2.1 System Overview

In this study, a data set containing information about a patient's characteristics who went through cryotherapy treatment was obtained. Binomial Logistic Regression will be explored as an approach for predicting the result of the cryotherapy treatment. The model will be comprised of building and training the model. The data set will be split into training and testing data. The first 75 samples in this data set will be used as training data and the last 15 samples will be used to evaluate the performance of the model. In the training data set, the data was made up of 37 samples that corresponded to patients whose cryotherapy treatment was not successful in healing warts and 38 patients whose cryotherapy treatment was successful in treating warts. This leaves the testing data to be comprised of 5 patients whose cryotherapy was not successful and 10 patients whose cryotherapy treatment was successful. The testing data is not balanced for validating the model, which might not give us enough information to validate the model. To solve this, the model will also be validated using 10 fold cross validation, where 90 percent of our data will be used for training the model and 10 percent of the data will be used to validate the model this method will provide us with more scores for validating the model which include sensitivity, AUC (Area Under the Curve), specificity and F1 measure.

2.2 Data set

The Cryotherapy data set from the UCI database contains records of ninety patients. The following table provides more information about the patients whose information was recorded:

Table 1: Cryotherapy Dataset Details

Features	Values
Age	15-67
Gender (Male to Female ratio)	47-43
Time elapsed before treatment	0-12
Number of warts	1-12
Area of warts (squared millimeters)	4-750
Type of wart	Common/ Plantar/ both
Result of Treatment	Healed or Not Healed

2.3 Machine Learning Method

Previous approaches that have been used in this data set have included Naive Bayes, Random Forest, Support Vector Machine, Artificial Neural Networks and k Nearest Neighbor algorithm. Previous approaches used these classifiers with all features and they also explored dimension reduction. The accuracies obtained by those methods ranged from 59 to 95 percent, sensitivities ranged from 82 to 96 percent, specificities ranged from 44 to 98 percent, AUC (Area Under the Curve) ranged from 58 to 95 percent and F1 Measure ranged from 64 to 94 percent. This study evaluates the performance of Binomial Logistic Regression without feature reduction and the model will be evaluated by measuring the classification accuracy of the model by dividing the dataset into 75 samples for training and 15 samples for evaluating the model. Another approach that will be used will be 10 fold cross validation where the dataset will be divided into using 90 percent of the dataset for training and 10 percent of the model will be used for validating the model. The scores that will be computed to validate the model are accuracy, sensitivity, specificity, AUC (Area Under the Curve), and F1 Measure.

When creating the model, the weights of the model needed to be initialized first. This was done using a built in function from numpy called numpy.random.rand, which generates a one or two dimensional array with random float values. In the Logistic Regression model, these weights will be used to calculate the dot product between the weights and each input. This dot product will be represented as a function g(x). Here x represents the features of each sample and the transpose of the weights is represented by θ^T.

$$\text{g(x)} = \theta^T x$$

The dot product will then be used as an input for the sigmoid function, shown below, here x represents the value that we get from g(x).:

Sigmoid Function

$$\sigma(x) = \frac{1}{1+e^(-x)}$$

This will return a float value between 0 and 1. To classify each sample, the float value will be compared with 0.5 which will be the decision threshold that will be used to classify each sample. The way that the classification will work is if the value returned by the sigmoid function is greater than or equal to 0.5, the sample is classified as 1 or it can be said that the treatment will be successful. Before the model is validated, the model will need to be optimized, that is, the weights of the model will have to be updated to minimize the error. In this study, the cost function, given below, is used to measure the error of our model. Here m represents the number of samples and h(x) represents the predicted value, that is h(x) = $\sigma(g(x))$.

Cost Function

$$\text{J}(\theta) = \frac{1}{m}\sum_{i=0}^{m}[-y^i \log(h(x^i)) - (1-y^i)\log(1-h(x^i))]$$

The optimization algorithm that is used in this study is gradient descent, shown below. (Step size is represented by μ.)

Gradient Descent

$$\text{J'}(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h(x^i) - y^i)x^i$$

$$\theta^i = \theta^(i-1) - \mu * J'(\theta)$$

The way that the model is updated is that the weights of the model will be updated using the gradient descent function for a given number of iterations. In this study, the model was updated using 20,000 iterations and a learning rate or step size of 0.0001.

3 Results and Observations

To evaluate the model, first the last 15 samples were used to evaluate the performance of the model. In this evaluation, of the fifteen samples thirteen were classified accordingly an two were misclassified, which leaves us with a classification accuracy of 86.67 percent.

The next method that was used to evaluate the model was 10 fold cross validation. The reason for using this model was because this method provides a less biased or less optimistic estimate of the performance of the model. The procedure that was followed in using this method was according to [5]:

1. Shuffling the dataset randomly
2. Splitting the dataset into 10 groups
3. For each unique group:
   a. Take one of the groups as the test data
   b. Use the other groups as training data
   c. Fit the model on the training data and evaluate it with the testing data
   d. Retain the evaluation scores
4. Summarize the skill of the model using the sample of model evaluation scores.

Using the k fold cross validation method, the scores of the model that were computed were accuracy, AUC (Area Under the Curve), sensitivity, specificity and F1 measure. The accuracy was calculated by counting the number of correctly classified samples and dividing that number by the total number of samples. For calculating AUC (Area Under the Curve), a function was imported from scikitlearn.metrics which uses the target scores for y and the predicted scores to calculate the Area Under the Curve, this score is used to see how well a parameter can distinguish between two diagnostic groups. Another evaluation score that was computed was F1 measure, which was computed by multiplying the sensitivity and precision by 2 and dividing that product by the sum of the precision and the sensitivity. The average scores that were obtained after multiple runs were:

Table 2: Results

Evaluation Measurements	**Values**
Sensitivity	0.6975
Specificity	0.8007
Accuracy	0.7334
AUC	0.7722
F1 Measure	0.7208

The results show that using Binomial Logistic Regression produced an Accuracy, Specificity, AUC, and F1 Measure that was higher than that of Artificial Neural Networks with all features,even though Articial Neural Network still achieved a higher Sensitivity, which was used in the previous study, but it was still outperformed by the other models that were used by Ali Cuvitoglu and Zerrin Isik.

4 Conclusion and Future Work

In this study, the Cryotherapy dataset published in the UCI KDD Archive by F. Khozeimeh et al, was analyzed for classifying the potential outcome of wart treatment. In a previous study conducted by Ali Cuvitoglu and Zerrin Isik, multiple machine learning approaches were studied with the best performing machine learning model being the Random Forest achieving an Accuracy of 95 percent, an F1 Measure of 94 percent, a Sensitivity of 88 percent, AUC (Area Under the Curve) of 95 percent, and a Specificity of 98 percent and Artificial Neural Networks producing the lowest scores achieving an Accuracy of 59 percent, an F1 Measure of 64 percent, an AUC (Area Under the Curve) of 58 percent and a Specificity of 44 percent, but still achieving a Sensitivity of 90 percent. In this study, Binomial Logistic Regression was explored as an approach for this classification task, since this was a binary classification task, for which Binomial Logistic Regression has been commonly used. With this approach, the results achieved were a Sensitivity of 69.75 percent, a Specificity of 80.07 percent, an Accuracy of 73.34 percent, an AUC of 77.22 percent, and an F1 Measure of 72.08 percent. While the approach explored in this study outperformed an Artificial Neural Network, it was still outperformed by most of the approaches explored previously. This could possibly be due to the fact that some features in the dataset are categorical, which would make a decision tree or a random forest a preferable approach. Another reason Logistic Regression did not perform as well as most models for this classification task is the likelihood that our data was not linearly separable.

For future work, this dataset will be explored using models that do not require that our data be linearly separable. Dimension reduction and feature selection will also be used as an approach on this dataset to see if they produce more optimal solutions.

References

[1] UCI KDD Archive. [Online]. Available: http://archive.ics.uci.edu/
[2] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, A. Khosravi, P. Layegh, and S. Nahavandi, ’An expert system for selecting wart treatment method,’ Computers in Biology and Medicine, vol. 81, pp. 167-175, 2/1/ 2017
[3] F. Khozeimeh, F. Jabbari Azad, Y. Mahboubi Oskouei, M. Jafari, S. Tehranian, R. Alizadehsani, et al., ’Intralesional immunotherapy compared to cryotherapy in the treatment of warts,’ International Journal of Dermatology, 2017, DOI: 10.1111/ijd.13535
[4] Ali Cuvitoglu and Zerrin Isik, "Evaluation Machine-Learning Approaches for Classification of Cryotherapy and Immunotherapy Datasets", *International Journal of Machine Learning and Computing Vol 8, No. 4 August 2018*
[5] Machine Learning Mastery [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/