Estimating Efficiency of Data Classification using Machine Learning Algorithm

Jeevan N
PG Scholar,
Department of MCA,
Dayananda Sagar College of Engineering,
Bengaluru, India
mailto:jeevan.n186@gmail.com

Dr.chandrika Murali
Assistant Professor,
Dayananda Sagar College of Engineering,
Bengaluru, India
mailto:chandrika-mcavtu@dayanandasagar.edu

***Abstract*— Machine learning, which can be summed up as teaching a computer to improve its predictions with experience, is a rapidly growing topic. With the help of recent rapid increases in computer storage capacity and computing power, we have recently been able to. In bioinformatics, machine learning techniques are commonly used.**

Due to the high cost and complexity of biological analyses, researchers have developed cutting-edge machine learning methods specifically for this field. In this research, we use a variety of classification methods on the iris flower dataset under the assumption that doing so will prove the efficacy of the machine learning technique.

Then we point to their performance evaluation. In this research paper, the performance among three selected kernel functions, polynomial, sigmoid, and RBF were evaluated using the iris flower dataset and the accuracy and F1 score values were estimated to select the best classifier for flower species classification.

Keywords:Machine Learning, Classification, Efficiency, Predictions, Accuracy.

I. Introduction

Machine learning is a branch of AI that enables computers to "self-learn" from training data and refine their results over time with no human intervention or supervision. Algorithms trained using Machine Learning may analyse data for patterns and utilise those patterns to make inferences. In a nutshell, Machine Learning uses experience to train its algorithms and models.

Traditional programming entails a computer engineer writing a set of instructions for the computer to follow in order to convert one set of data into another. The IF-THEN structure is widely used in programming instructions; if a condition is fulfilled, the programme will execute the corresponding "then" statement; testing can be used to refine the findings [3, 7].

In contrast, machine learning is a sort of automation that enables computers to learn how to solve issues and carry out specific tasks with little to no guidance from a human operator. In contrast, machine learning is an automated method that enables robots to solve issues with minimal or no human intervention.

You may train machine learning algorithms to conduct computations, processes data, and automatically recognise patterns without explicitly programming them to do so.A supervised learning algorithm is one that learns from examples given to it. Algorithms for unsupervised learning find patterns and correlations in data that has not been labelled. In this scenario, models are fed data to process but aren't told what the desired results are; as a result, they have to make inferences based on evidence alone, without any direct guidance.. The models are not trained with the "correct answer," so they must discover the patterns on their own; testing was performed to validate the results [9].

II. II. LITERATURE SURVEY

Boaz LernerandAmit David [1] implemented a support vector machine that categorizes current cytogenetic signals from fluorescence in-situ hybridization photographs with respect to the evaluation of genetic syndromes. The research works implement the concept of minimizing structural threats of SVM in finding the most suitable kernel structure and classifier parameters. The results reveal that SVM performs correctly in identifying FISH signals when compared to other cutting-edgeclassifiers for machine learning, demonstrating the possibilities of the SVM-based framework for genetic analysis.

Saeed Shariati. used fuzzy NN with SVM and ANN. [2] to forecast and identify thyroid illness and hepatitis. Furthermore, through disease analysis, they were able to determine the type and stage of the disease, which has five groups for thyroid disorders,includes thyroid gland absence, subclinical hypothyroidism, thyrotoxicosis, and hypothyroidism,as well as six categories of hepatitis disease, such as hepatitis, non-hepatitis, hepatitis B (two stages), and hepatitis C (two stages).with 99% accuracy due to thyroid illness and 98% accuracy for hepatitis

An effective Parkinson's disease analysis system utilising LSTSVM and Particle Swarm Optimization PSO was provided by D. Tomar et al. [4]. PSO is employed for parameter and characteristic optimisation.Regarding precision, specificity, and sensitivity, the performance of the suggested process is contrasted with many other present procedures. Empirical findings demonstrate that the suggested Parkinson's disease evaluation approach is 98% accurate when compared to other existing procedures.

In order to anticipate cardiac disease, Yan Zhang et al. [5] employed SVM with RBF kernel function,This has statistical learning theory at its foundation. The Grid Search method is used to select excellent kernel function parameters and a fine basis function to refine the criteria to achieve probably the highest classification accuracy.

Fatemeh Saiti proposed using SVM and Probabilistic NN. [10] for the classification of the two thyroid disorders from the thyroid disorders database, hypothyroidism and hyperthyroidism.To deal with redundant and irrelevant information, these algorithms frequently use effective classification techniques. A clever and potent method for selecting fair subsets of factors that provide more accurate predictions was tested by genetic algorithms. Support Vector Machines (SVMs), which are particular types of brain activities, were employed by D. Vassis. [11] reviewed the use of neural networks in automated medical prognostication in great detail.In many circumstances, neural programmes can also accurately forecast symptoms and diseases during evaluation, SVMs, on the other hand, are gradually being used in clinical prognosis because of their distinctive categorising properties.

III. METHOD USED AND RESULT

This code makes use of the Iris flower dataset. This dataset has three classes, each with 50 instances, where each class represents a different variety of iris plant.

To evaluate the effectiveness of the two chosen kernel functions on the same data file, the code additionally determines accuracy and f1 score.fromsklearn.

In the first step iris dataset is loaded

importsvm, datasets

importsklearn.model_selection as model_selection fromsklearn.metrics

import accuracy_score fromsklearn.metrics import f1_score

iris = datasets.load_iris()

x = iris.data[:, :3]

y = iris.target

then training and testing data were prepared

Model_selection.trn_tst_split(x, y, train_size=0.80, test_size=0.20, random_state=101) splits a test into two halves. deliver the pursuing values x_trn, x_tst, as well as y_trn, y_tst

Next support vector machine kernel functions were used

rbf = SVM.svc (gamma of 0.5, C of 0.1, and rbf kernel).x_trn and y_trn

poly is same to SVM.svc(poly kernel, gamma=0.5, degree=3, and C=0.1).x_trn and y_trn

sig=SVM.svc(degree=3, gamma=0.5, C=0.1, kernel='sigmoid').X_trn and Y_trn

next prediction done using different kernel functions

poly_predict = poly.pred(x_tst)

rbf_predict = rbf.pred(x_tst)

sig_predict = sig.pred(x_tst)

then f1 and accuracy scores were calculated

poly_acc = (y_tst, poly_predict) acc_score Poly_f1 is calculated as f1_score(y_test,Poly_predict, avg='weighted')

print('Accuracy of the Polynomial kernel: ', "%.2f" % (poly_acc*100))

print('F1 Polynomial Kernel Accuracy: ', "%.2f" % (poly_f1*100))

y_test, rbf_predict, rbf_acc = acc_score y_test, rbf_predict, avg='weighted'; rbf_f1 = f1_score

print('RBF Kernel Accuracy: ', "%.2f" % (rbf_acc*100)) print('F1 RBF Kernel Accuracy: ', "%.2f" % (rbf_f1*100))

sig_acc = acc_score(y_test, sig_predict) sig_f1 = f1_score(y_test, sig_predict, avg='weighted')

print(“Accuracy of the Sigmoid kernel:=”,sig_acc*100) print(“F1 Score of Sigmoid kernel:=”,sig_f1*100)

Accuracy of the Polynomial Kernel: 96.67

F1 Polynomial Kernel Accuracy: 96.62

RBF Kernel Accuracy: 93.33

F1 RBF Kernel Accuracy: 93.33

Accuracy of Sigmoid kernel: 26.666666666666668

F1 Score of Sigmoid kernel: 11.228070175438596

IV. RESULTS AND FINDINGS

In this study, a machine learning algorithm and various kernel function types with support vector machine classification were used to train a machine learning model by splitting the iris flower dataset into training and testing data. The classification was then assessed using the iris dataset, in which the accuracy and Values

F1 scores were estimated to select the best classifier for classifying different flower species.

F1 scores were estimated to select the best classifier for classifying different flower species.

Sigmoid Kernal Function whose accuracy values are 26.66 and F1 score is11.22.

IV. REFERENCES

[1] Amit David And Boaz Lerner,” Pattern Classification Using A Support Vector Machine For Genetic Disease Diagnosis” Proceedings.2004 23rd Ieee Convention Of Electrical And Electronics Engineers In Israel, Pp 289~292.

[2] Saeed Shariati, Mahdi MotavalliHaghighi ,”Comparison Of Anfis Neural Network With Several Other AnnsAndSupport Vector Machine For Diagnosing Hepatitis And Thyroid Diseases”, International Conference On Computer Information Systems And Industrial Management Applications (Cisim), 2010,Pp 596~599.

[3] .Udupa,Pradeep."AnIntegrated Methodology For Testing Source Code By Using Multi Constraint Reduction, Test Suite Prioritization And Prioritized Parallelization",Ijet,Vol.10,No2,Pp.221-225.

[4] D. Tomar; B. R. Prasad; S. Agarwal,” An Efficient Parkinson Disease Diagnosis System Based On Lea St SquarEs Twin Sup Port Vector Machine And Part IcleSwarm Optimization”9thInternational Conference On Industrial And Information Systems(Iciis),2014,Pp 1-6.

[5] Yan Zhang, Fugui Liu, Zhigang Zhao, Dandan Li, Xiaoyan Zhou, Jingyuan Wang,” Studies On Application of Support Vector Machine In Diagnose of Coronary Heart Disease” Sixth International Conference On Electromagnetic FieldProblems And Applications (Icef).2012,Pp.1-4.

[6] Bo Pang, David Zhang, Naimin Li, And Kuanquan Wang,” Computerized Tongue

DiagnosisBased On Bayesian Networks",IeeeTransactions On Biomedical Engineering, No.10,Vol. 51, October 2004, Pp 1803-1810.

[7].Udup,Pradeep."ProductiveMechanism ToValidate Program Segment By Using Test Case ReductionAndTestSuitePrioritization",Ijeat 2019, Vol. 10, No 2, Pp. 221–225.

[8] JavadSalimiSartakhti, MohammadHosseinZangooei, KouroshMozafari" HepatitisDiseaseDiagnosis Using A Novel Hybrid Method Based On Support Vector Machine And Simulated Annealing (Svm-Sa)",ComputerMethods And Programs In Biomedicine,2 0 1 2,Vol. 108,Pp. 570–579.

[9] Udupa, Pradeep. "Application of Artificial IntelligenceFor University InformationSystem."Engineering Applications of Artificial Intelligence 114 (2022): 105038.

[10] FatemehSaiti, AfsanehAlaviNaini, Mahdi AliyariShoorehdeli, Mohammad Teshnehlab," Thyroid Disease Diagnosis Based On Genetic Algorithms Using Pnn And Svm",Ieee 3 rdInternational Conference On Bioinformatics And Biomedical Engineering, Icbbe 2009, Pp 1-4.

[11] D.Vassis,B.A. Kampouraki, P. Belsis, V. Zafeiris, N. Vassilas, E. Galiotou, N. N. Karanikolas, K. Fragos, V. G.Kaburlasos, S. E. Papadakis, V. Tsoukalas, And C. Skourlas," Using Neural Networks And Svms For Automatic Medical Diagnosis: A Comprehensive Review", International Conference On Integrated Information (Icininfo 2014) AipConf. Proc. 1644, 32-36 (2015);Doi:10.1063/1.4907814.