Predictive Modeling Using Neural Networks and Support Vector Machines: A Survey

Deeptesha Satpute, Karan Kanitkar, Kedar Bartake, Pranali Lokhande, Dr. K. Rajeswari
Department Of Computer Engineering,
Pimpri Chinchwad College of Engineering,
Savitribai Phule Pune University, India.
deepteshasatpute29@gmail.com, smartkenny27@gmail.com, kedarbartake98@gmail.com,
pranali.lokhande2@gmail.com, raji.pccoe@gmail.com

***Abstract*— Machine Learning is the science of enabling computers to learn phenomena which is very similar to the human learning process but is not limited to its limitations. Machine Learning algorithms are designed to process and detect patterns in voluminous data. Machine learning is believed by many to be the most efficient and promising way to reach towards the goal of truly intelligent computers. Artificial neural networks are structures that were designed to use computers to mimic the working of the human brain. They are very popular and hugely powerful learning algorithms which are capable of performing highly complex mathematical operations and pattern recognition tasks. Another popular machine learning algorithm used for classification tasks is Support Vector Machines (SVM).They can be easily modified using different kernel functions and thus are robust and versatile. Cross-validation techniques are used to improve the accuracy of predictive models. Ensemble Learning methods are used to combine many models to improve over each of them.**

***Index Terms*—Machine Learning, Artificial Neural Networks, Support Vector Machines, Cross-Validation, Ensemble Learning, Loss Functions, Activation functions.**

I. INTRODUCTION

Machine learning is the ability of computers to perform tasks without being explicitly programmed. Formally speaking, 'A system Performing a task T, with performance measure P, and Learning experience E, is said to be a machine learning system if its performance P on task T, improves with experience E. For Example: Google's Search Engine, Driver-less Cars, Recommender Systems, Robotics, Handwritten Digit Classification, etc. Further sections of the paper describe types of Machine Learning algorithms, and then explores Classification, two algorithms used for classification- Artificial Neural networks and Support Vector Machines. Further, the methods to improve accuracy of the models generated are explored- namely cross-validation and ensemble learning techniques.

II. CLASSIFICATION OF MACHINE LEARNING ALGORITHMS

Machine Learning Algorithms can be classified into two main categories:

- Supervised Learning Algorithms
- Unsupervised Learning Algorithms

A. Supervised Machine Learning

In Supervised Machine Learning, we are given a data set and already know the 'right answers' to each of the training examples provided in the data set. We also have an idea about the relationship between the inputs and the outputs. Supervised Learning problems are divided into two main categories, namely, regression problems and classification problems [9]. In a regression problem, the output is a continuous function, i.e., we are mapping input variables to a continuous function. For example, given the data about the size of houses on the real estate market, trying to predict their price as a function of size.

In a classification problem, we try to predict the results in a discrete output. In other words, we try to map input variables into discrete categories. For example, trying to predict whether an image is that of a cat, a dog, or a ship. Here, we classify the input images into three different categories. Polycystic Ovary Syndrome [21] is one of the applications of supervised machine learning techniques.

B. Unsupervised Machine Learning

Unsupervised Machine Learning Algorithms are used primarily, to explore the inherent structure of the data. The data provided to these algorithms does not have the 'right answers' to it. Cluster analysis is the most common unsupervised learning method which can be used for exploratory data analysis for finding hidden patterns of grouping in data. Hierarchical clustering, K-means clustering and Gaussian mixture models are some of its types. One of the applications of unsupervised machine learning is web intrusion detection systems [16].

III. CLASSIFICATION

Classification is a supervised machine learning task in which the output is discrete i.e. the output is one label from a finite set of labels [3]. For e.g. in the case of handwritten digit recognition system, the input is an image of a handwritten digit and the output is one from the set of numbers, {0-9}. Algorithms used for classification are Decision trees, Artificial Neural Networks, Advanced Support Vector Machines, Logistic Regression, Naïve Bayes etc. [3].

IV. ARTIFICIAL NEURAL NETWORKS

A. *Introduction*

Artificial neural networks are structures that were designed to use computers to mimic the working of the human brain [17]. Artificial neural networks are very popular and hugely powerful learning algorithms which are capable of performing highly complex mathematical and pattern recognition tasks [18]. These systems are so complex that even prominent machine learning professionals and mathematicians are unable to decipher what exactly happens in them.

B. *Structure of Neuron*

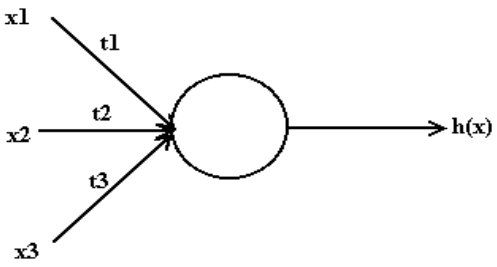

Fig 1. Artificial Neuron [15]

The above diagram is the basic structure of an artificial neuron. The inputs x1, x2, x3 are multiplied by weights t1, t2, t3 and are added. The result is passed as an input to the activation function [13]. The output of the activation function is the output (O) of the neuron.

$$O=activation\ function\ (x_1t_1+x_2t_2+x_3t_3) \quad (1)$$

Eq. 1 represents the fundamental operation that occurs in the neuron in terms of the inputs and weights.

C. *Structure of Artificial Neural Network*

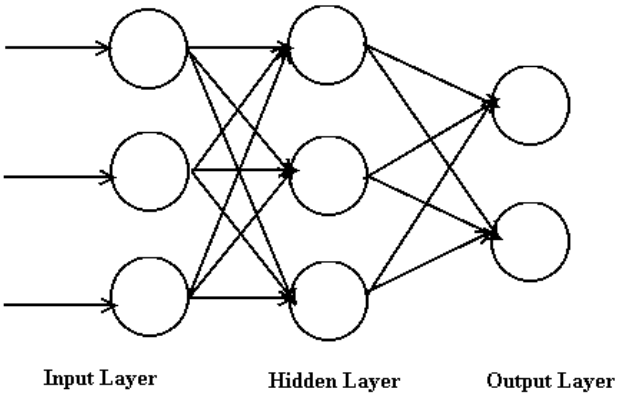

Fig 2. Artificial Neural Network [15]

Fig. 2 represents a neural network which consists of several neurons. Artificial Neural Network consists of one input layer, one or more hidden layers and, one output layer [15]. Due to the presence of one or more hidden layers, Neural Networks are capable of processing or predicting highly complex functions. Also due to this, they are capable of designing their own features.

D. *Types of Artificial Neural Networks*

According to the requirement of application, a specific type of neural network is selected. Some of them include:

a) Convolutional Neural Network [5]: These are used for image processing and classification tasks.
b) Recurrent Neural Network [5]: These are used for speech recognition tasks and classification tasks.
c) Radial Basis Function Neural Network [20]: These are used to combine parametric statistical distribution model and non-parametric linear perceptron algorithm sequentially. It is an alternative technique to Back Propagation Neural Network.

E. *Activation Functions* [14] [15]

Activation functions are responsible for processing the input and providing the output. The choice of activation functions greatly impacts the quality of output in terms of accuracy. Neural networks can be designed using different activation functions.

F. *Types of Activation Functions* [13] [14]

Selection of activation function greatly impacts the result of Artificial Neural Network. Different types of activation functions are:

a) Sigmoid Function:
The equation of the sigmoid function is

$$f_{sig}(x) = \frac{1}{1+e^{-x}} \quad (2)$$

b) Hyperbolic Tangent Function (tanh)

$$\tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \quad (3)$$

c) Rectified Linear Unit (ReLU)

$$f(z) = \max(0, z) \quad (4)$$

d) Leaky ReLU

$$f(z) = \begin{cases} 0, & if\ z > 0 \\ az, & otherwise \end{cases} \quad (5)$$

$$note : if\ a \leq 1$$
$$f(z) = \max(z, az) \quad (6)$$

e) Softmax Function

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad for\ j = 1, \ldots\ldots, K \quad (7)$$

The sigmoid function (Eq. 2) and the tanh function (Eq. 3) are advantageous as their outputs can be used to calculate their derivatives. This property is very useful in the learning process of neural networks and saves a lot of computations. The Softmax function (Eq. 7), ReLU (Eq. 4), Leaky ReLU (Eq. 5) are used in convolutional neural networks in computer vision and speech recognition problems.

G. *Loss Functions.*

Loss functions calculate the error by which the output of the hypothesis differs from the training data. Thus, they are also the optimization objective of the Machine learning process. The lower the value of the loss, the better the hypothesis fits to the data [18]. These functions are also the metrics for calculating the generalization error of the predictive model. Thus, they serve as an important performance metric in predictive modeling.

H. *Types of Loss Functions*

Some examples of loss functions are:

a) Softmax Loss [5]
b) LambdaRank Loss [5]
c) Sum of Squares of Errors (SSE) [18]
d) Mean Square Error (MSE) [18]
e) Mean Absolute Error (MAE) [18]

The pairings of activation functions and loss functions are important factors that affect the performance of Neural Networks.

I. *Applications*

Some of the applications of Artificial Neural Network are Detection of Parkinson's disease [23] [10], Web crawlers [22], ECG Classification [19], Stock price forecasting [24], Medical image recognition [25].

V. SUPPORT VECTOR MACHINE

Support vector machine is another supervised learning algorithm used for classification, regression and outlier analysis. Support vector machines are used to construct a hyper plane in infinite dimensional space [4]. It is a large margin classifier. The advantage of support vector machine is that is effective in high dimensional spaces. The disadvantage of support vector machine is that if there is a large dataset then the training time is high.

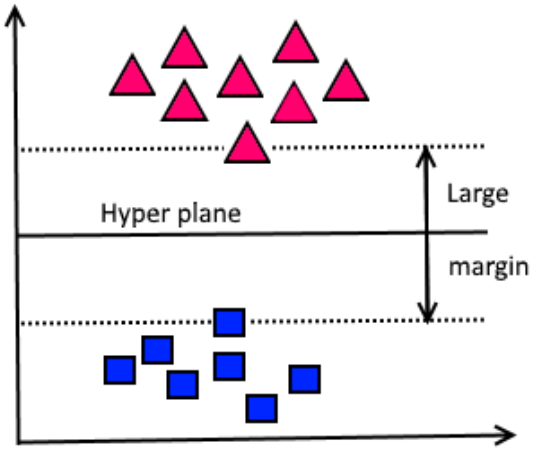

Fig 3. Large margin classifier using support vector machine [6]

In Fig. 3, the different shapes denote data points from different classes. And the hyperplane denotes the decision boundary obtained by the algorithm.

A. *Kernels* [3]

Kernels are used to define extra new features using the similarity functions and landmarks for training more complex non-linear classifiers.

Types of kernels are [3]:

1) *Radial Basis Function*

$$k(x, x') = \exp\left(\frac{-\|x - x'\|^2}{2\sigma^2}\right) \quad (8)$$

2) *Gaussian Function*

$$k(x, x') = exp\left(\frac{-\|x - x'\|}{\sigma^2}\right) \quad (9)$$

3) *Laplace Function*

$$k(x, x') = exp\left(\frac{-\|x - x'\|}{\sigma}\right) \quad (10)$$

B. *Applications*

Some of the applications of support vector machine are Ultrasonic flaw detection [6], Fraud detection [2].

VI. ENSEMBLE LEARNING

Ensemble learning techniques combine multiple classifiers into one better classifier. This leads to the better accuracy than

each of the constituent classifiers. The resulting model is better in performance than each of the constituent models [11]. Prominent algorithms in ensemble learning are AdaBoost [11], XGBoost and Random Forest [11]. The RandomForest algorithm, which is an ensemble of many decision trees, is an example of the use of such techniques. Prominent ensemble learning methods are Bagging and Boosting.

A. *Bagging* [1]

Bootstrap Aggregating or Bagging, is a technique which uses small changes in training data to generate very diverse classifiers which are then combined to produce a single classifier with low test error. RandomForest algorithm uses bagging method.

B. *Boosting* [1]

A Boosting is an ensemble learning method for combining several weak learners into one strong learners. A weak learner is one that has an accuracy which is only slightly greater than a classifier that outputs class labels randomly. A strong learner is a classifier whose accuracy is markedly greater than that of the random output classifier. The AdaBoost algorithm is an example of the Boosting approach.

VII. CROSS VALIDATION

Cross-validation techniques are used to make models that generalize well. i.e., they are not susceptible to the trends in training data. Therefore, they have lower generalization error [8]. They divide the training data into training and testing splits leaving out the test data and training on the remaining data at each iteration. After completing the iterations, the learned parameters are averaged. Thus, the model has a lower error value.

A. *Cross validation techniques*

Cross validation techniques are classified into two main classes:

1) *Hold out cross validation* [8]

In hold out validation, the data is split into two parts one of which is used for training and the other is used for training and other is used for testing. The model is trained over the training dataset and tested using the test data. The dataset used for testing is called the holdout set. The dataset used for testing is called the holdout set. The advantage of this approach is that it takes less time to train than other methods. The disadvantage is that it does not use the entire available data for the purpose of training the model and thus is non-exhaustive.

2) *k-fold cross validation*

In k-fold cross validation, the available data is divided into K equal sets further, at each iteration , one of them is used as test data while the remaining (k-1) sets together are used for training [7] [12]. Thus, at every iteration, one data set is held out and is used for testing. At the end of all iterations, the model parameters are averaged and the error is thus reduced. In this process, the model is trained using the entire available data and thus is exhaustive. Thus, in cases where the data available is limited, k-fold cross validation can result higher accuracy [4]. Adjusting values of k for higher accuracy using trial and error is the pitfall of this approach.

VIII. CONCLUSION

Machine Learning is used in many real world problems. There are many innovative products and applications that use machine learning algorithms like Google DeepDream, Translator etc. Supervised Machine Learning Algorithms like Artificial Neural Networks, Support Vector Machines, etc. are used for classification purposes. Artificial Neural Networks are very flexible as they can be customized by using different activation and loss functions. Support Vector Machines can be customized by using different kernel functions. Cross Validation techniques are used to improve the prediction accuracy and generalization. Furthermore, ensemble learning methods can combine the capabilities of multiple learning models to produce a single model that can be much accurate than rest of the models.

REFERENCES

[1] Jianchang, “A Case Study on Bagging, Boosting, and Basic Ensembles of Neural Networks for OCR “, 1998, California.

[2] Abiodun Modupe, Oludayo, O. Olugbara Sunday, O. Ojo, “Exploring Support Vector Machines and Random Forests to Detect Advanced Fee Fraud Activities on Internet”, 11th IEEE International Conference on Data Mining Workshops,2011.

[3] Dr.P.Arumugam, P.Jose, “Recent Advances on Kernel Fuzzy Support Vector Machine Model for Supervised Learning “, International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2015.

[4] Zahra Nematzadeh, Roliana Ibrahim, Ali Selamat, “Comparative Studies on Breast Cancer Classifications with K-Fold Cross Validations Using Machine Learning Techniques “, IEEE International Conference on Data Mining, 2014.

[5] Chunyan Xu, Canyi Lu, Xiaodan Liang, Junbin Gao, Wei Zheng, Tianjiang Wang, Shuicheng Yan, “Multi-loss Regularized Deep Neural Network”, IEEE transactions on circuits and systems for video technology, vol. 26, no. 12, December 2016.

[6] Kushal Virupakshappa, Erdal Oruklu, “Ultrasonic Flaw Detection using Support Vector Machine Classification “.

[7] Ozsel Kilinc, Ismail Uysal, “Source-aware Partitioning for Robust Cross-validation”, IEEE 14th International Conference on Machine Learning and Applications, 2015.

[8] Sanjay Yadav,Sanyam Shukla, “Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification”, Computing Conference, IEEE 6th International Conference on Advanced Computing,2016.

[9] Chuncao Zhang ,Guoli Zhu, Lan Yue,“An advanced detection approach based on support vector machine during tunnelling “,2016.

[10] Decho Surangsrirat, Chusak Thanawattano, Ronachai Pongthornseri, Songphon Dumnin, Chanawat Anan, and Roongroj Bhidayasiri, “Support Vector Machine Classification of Parkinson’s Disease and Essential Tremor Subjects based on Temporal Fluctuation “,2016.

[11] Ahmet Tartar , Aydin Akan ,”Ensemble Learning Approaches to Classification of Pulmonary Nodules”, 2016.

[12] Michal Vasinek, Jan Platos,Vaclav Snasel ,“Limitations on Low Variance k-Fold Cross Validation in Learning Set of Rules Inducers”, International Conference on Intelligent Networking and Collaborative Systems,2016.

[13] Rodrigo G. F. Soares ,Emeson J. S. Pereira," On the performance of pairings of activation and loss functions in neural networks",2016.
[14] Andrich B. van Wyk ,Andries P. Engelbrecht "Analysis of Activation Functions for Particle Swarm Optimised Feedforward Neural Network",2016.
[15] Yin Liu, Hariharasudhan Venkataraman, Zisheng Zhang and Keshab K. Parhi," Machine Learning Classifiers using Stochastic Logic",2016.
[16] Truong Son Pham, Tuan Hao Hoang,Van Canh Vu," Machine learning techniques for web intrusion detection – a comparison " ,Eighth International Conference on Knowledge and Systems Engineering (KSE),2016.
[17] Honey Mehta, Sanjay Singla, Aarti Mahajan ,"Optical Character Recognition (OCR) System for Roman Script & English Language using Artificial Neural Network (ANN) Classifier ",2016.
[18] Danilin S.N., Makarov M.V., Shchanikov S.A. "Design of Artificial Neural Networks with a Specified Quality of Functioning", International Conference on Engineering and Telecommunication, 2014.
[19] Shweta H. Jambukia , Vipul K. Dabhi , Harshadkumar B. Prajapati ,"Classification of ECG signals using Machine Learning Techniques: A Survey", International Conference on Advances in Computer Engineering and Applications (ICACEA),2015.
[20] P. Jeatrakul and K.W. Wong," Comparing the Performance of Different Neural Networks for Binary Classification Problems", Eighth International Symposium on Natural Language Processing, 2009.
[21] Palak Mehrotra, Jyotirmoy Chatterjee, Chandan Chakraborty, Biswanath Ghoshdastidar, Sudarshan Ghoshdastidar," Automated Screening of Polycystic Ovary Syndrome using Machine Learning Techniques".
[22] Abhiraj Darshakar, "Crawler intelligence with Machine Learning and Data Mining integration", International Conference on Pervasive Computing (ICPC), 2015.
[23] Pawalai Kraipeerapun, Somkid Amornsamankul, "Using Stacked Generalization and Complementary Neural Networks to Predict Parkinson's Disease", 11th International Conference on Natural Computation (ICNC), 2015.
[24] Nguyen Lu Dang Khoa, Kazutoshi Sakakibara and Ikuko Nishikawa," Stock Price Forecasting using Back Propagation Neural Networks with Time and Profit Based Adjusted Weight Factors", SICE-ICASE International Joint Conference 2006 Oct. 18-2 1, Busan, Korea.
[25] Tadashi Kondo, Junji Ueno, Schoichiro Takao," Hybrid Feedback GMDH-type Neural Network Using Principal Component-Regression Analysis and Its Application to Medical Image Recognition of Heart Regions", SCIS&ISIS 2014, Kitakyushu, Japan, December 3-6, 2014.