

# A Review on Classification Techniques in Machine Learning

R. Vijaya Kumar Reddy<sup>1</sup>, Dr. U. Ravi Babu<sup>2</sup>

<sup>1</sup>Research Scholar, Dept. of CSE, Acharya Nagarjuna University, Guntur, (India)

<sup>2</sup>Principal, DRK College of Engineering & Technology, Hyderabad, (India)

## ABSTRACT:

A Classification is a method of predicting similar information from the value of a categorical target or categorical class variable. It is a useful technique for any type of statistical data. These algorithms are used for various purposes like image classification, Predictive modeling, data mining technique etc. The main purpose of supervised learning is to build a simple and unambiguous model of the allocation of class labels in terms of predictor features. The classifiers are then used to classify class labels of the testing instances where the values of the predictor features are known, to the value of the class label which is unknown. In this paper we illustrate various classification techniques used in supervised machine learning.

**Keywords:** Classification, supervised, machine learning, pattern recognition.

## I.INTRODUCTION

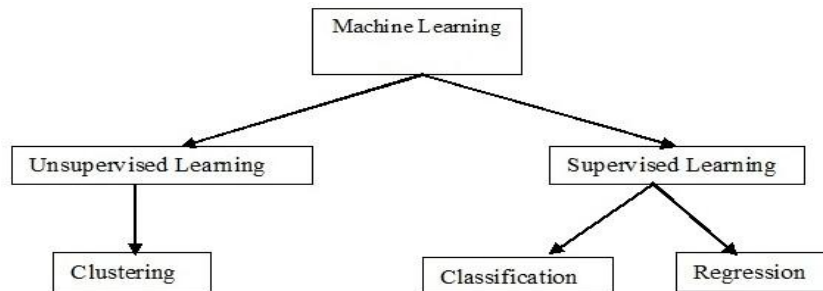
Machine learning approach is a technique used to teach machines how to handle the data more efficiently and get result More accuracy. In Some cases after viewing the data, we cannot understand the pattern or extract information from the data. In such case, we apply machine learning techniques for predicate the data [1]. Large quantity of datasets are available from different sources, there is a demand for machine learning. Many industries from medicine to military are applying machine learning to extract relevant information from the available datasets. The main purpose of machine learning is to learn from the existing data. Large set of algorithms are design how to make machines learn by themselves [2] [3]. Many mathematicians and programmers apply several approaches to find the solution of this problem. Some of them are demonstrated in Fig. 1. All the supervised learning Classification techniques of machine learning are explained in Section 2. Section 3 concludes this paper.

## II.TYPES OF LEARNING

A machine learning system learns from past experiences to improve the performances of intelligent application programs. Machine learning system is category into two types

- Supervised Learning
- Unsupervised Learning

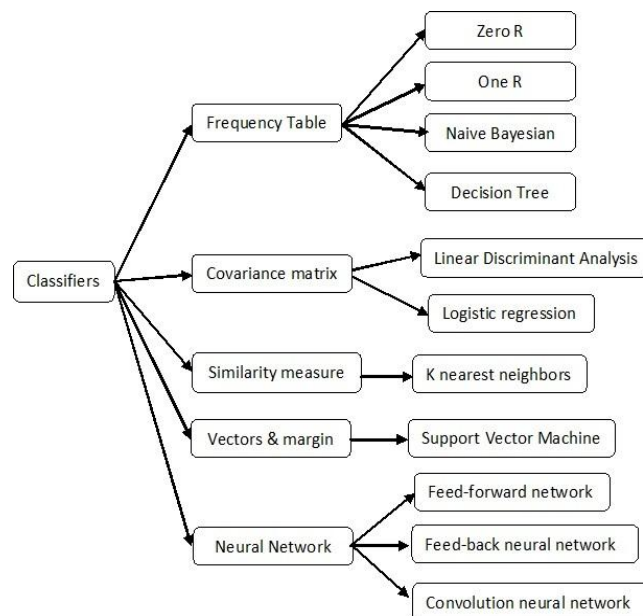
Supervised learning builds the learning model that effectively “learns” how to estimate from training data of given example. Unsupervised Learning builds a model based upon "unlabeled" data and to estimate key features of the data and characterized them without any prior knowledge of data.



**Fig-1** Types of Machine Learning

**2.1 Supervised Learning**

In this paper we describe various classification techniques in supervised learning. In supervised learning we divide the entire dataset into two parts one for training where the classifier learn form that data and remaining data is used for testing accuracy of the classifier. Once it is done then we can used to test new data for predicate the future information from these supervised learning classifiers. Supervised learning classifiers are classified in five main groups of classification algorithms base on Frequency Table, Covariance matrix, Similarity measure, Vectors & margin and Neural Network. From this group of classification we have different classification algorithms.



**Fig-2** Different classification algorithms in Supervised Learning

### 2.1.1. ZeroR

ZeroR is the simplest classification method which depends on the target data and ignores remaining all predictors. ZeroR classifier simply predicts the majority category labels. Although there is no predictability power in ZeroR, it is useful for determining a baseline performance as a standard for other classification methods [4].

#### Algorithm

Construct a frequency table for the target and select its most frequent value.

### 2.1.2 OneR

OneR is also known as One Rule, which simple classification algorithm used to generates one rule for each predictor in the data but not much accurate. It select only one of the best predictor from frequency table for predicate the target, which as the smallest total error using OneR algorithm. It also as slightly less accurate than state-of-the-art classification algorithms [4].

#### Algorithm

For each predictor,

For each value of that predictor, make a rule as follows;

Count how often each value of target (class) appears

Find the most frequent table

Make the rule assign that class to this value of the predictor

Calculate the total error of the rules of each predictor

Choose the predictor with the smallest total error.

### 2.1.3 Naive Bayesian

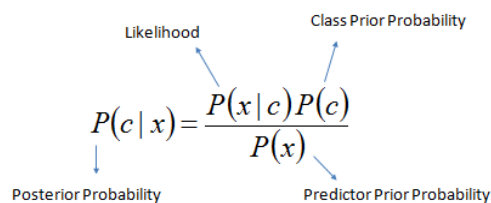
The Bayes' theorem is depending on Naive Bayesian classifier with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Regardless of its simplicity, the Naive Bayesian classifier it often outperforms more sophisticated classification methods and often does surprisingly well and is widely used.

#### Algorithm

Bayes theorem provides a way of calculating the posterior probability,  $P(c|x)$ , from  $P(c)$ ,  $P(x)$ , and  $P(x|c)$ .

Naive Bayes classifier assumes that the effect of the value of a predictor ( $x$ ) on a given class ( $c$ ) is independent of the values of other predictors.

This assumption is called class conditional independence.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$


$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

#### 2.1.4 Decision Tree

Decision tree builds classification models in the form of a hierarchical structure. Decision tree is developed through step by step incremental process of breaking down the dataset into smaller and smaller. At final process it generates a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The root node in a tree which corresponds to the best predictor from given datasets.

Decision trees classifier can use for both categorical and numerical data [6].

##### **Algorithm**

1. The root of the tree is select from the attribute of the dataset by using the concept of information gain.
2. Split the training dataset into subsets. And these subsets prepared in such a way that each subset contains data with the same value for an attribute.
3. Continue the process of step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

##### **Entropy**

A decision tree is constructed base on top-down approach from a root node and involves partition of data into subsets that contain instances with similar values upon to leaf nodes. The main algorithm for construction decision trees called ID3 which employs a top-down approach, ID3 algorithm uses entropy to calculate the similarity in sample. If the sample is totally similar then the entropy is zero and if the sample is not similar then divided its entropy of one.

To generate a decision tree, we need to determine two types of entropy using frequency tables as follows:

- Entropy with the frequency table of one attribute.
- Entropy with the frequency table having two attributes.

##### **Information Gain**

The information gain is based on the reduce in entropy after a dataset is divided onto an attribute. Developing a decision tree is all about finding attribute that returns the highest information gain i.e homogeneous

#### 2.1.5 Linear Discriminant Analysis

Covariance Matrix method is used for Linear Discriminant Analysis (LDA). More complex methods like mathematically method and often produces models is good to generate accuracy [5].

##### **Algorithm**

Linear combination of variables (predictors) concept is used in LDA which is based upon searching for a that best separates two classes (targets).

#### 2.1.6 Logistic Regression

Logistic regression predicts the probability of an outcome that can only have Boolean values. The prediction is done on both numerical and categorical. A linear regression is not suitable for predict the value of a binary variable for two reasons

A linear regression cannot predicate the values with in acceptable range. Since the dichotomous experiments can only have one of two possible values for each experiment, the residuals will not be normally distributed about the predicted line. But logistic regressions produce a logistic curve, which is limited to values between 0 and 1. Logistic regression is similar to a linear regression, but the curve is constructed using the natural logarithm rather than the probability. Moreover, equal variances in each group or the predictors will have unusually distribution [5].

### **2.1.7 K Nearest Neighbors**

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition based on their nearest neighbors and it should odd number .It uses the distance factors like Euclidean, Manhattan, Minkowski etc [6].

#### **Algorithm**

K nearest neighbors measured by a distance function like Euclidean, Manhattan, Minkowski etc. Most of the case we taken K as odd number only for making the decision.

If  $K = 1$ , then simply assigned to the class of its nearest neighbor.

Or

If K is odd number, then assigned to the class of maximum votes of its nearest neighbor.

### **2.1.8 Support Vector Machine**

A Support Vector Machine (SVM) classifies information by finding the maximized hyper plane that used as margin between the two classes [6].

#### **Algorithm**

1. Generate different hyper-plane and then identify the right hyper-plane.
2. Optimize the hyper plane with maximize margin between the classes
3. The kernel trick for non –linear hyper-plane used in SVM technique for misclassifications of linear hyper-plane.
3. For high dimensional space where we reformulate problem so that data is mapped implicitly to this space.

### **2.1.9 Feed-forward neural network**

A feed-forward network is a non- repetitive network which travels in one direction. It contains input, output and hidden layers. Elements are passed in input layer for processing data to calculations. There will have link between input, hidden and output layers and each input will have some weight. These weights are processed and make computation based upon the weights of inputs. It will be calculated and forward ad input to other hidden layers and it is counties until it reached to output. A threshold function is used to quantify the output of a neuron in the output layer [7][8].

### **2.1.10 Feed-back neural network**

A feed-back network has back propagation of feed-back paths which can travel in both directions using repetitive loops. All possible connections between neurons are allowed. Since repetitive are present in this type

of network, it becomes a non-linear dynamic system which changes continuously until it reaches a state of equilibrium. Feed-back networks predicted output of the neural network is compared with the actual output. Based on the error, the parameters are changed, and then fed into the neural network again to optimization problems where the network will be get best arrangement of interconnected neurons [7][8].

### 2.1.11 Convolutional neural network

Convolutional Neural Networks are very similar to regular Neural Networks .The neurons present in this network are have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. Image classification is processes of accepting an input image and generating output class (a cat, dog, etc) or a probability of classes that best match to the image [9].

Name of the Classification	Advantage	Disadvantage
<b>ZeroR</b>	Provided standard for other classification methods	Depends only on target data
<b>OneR</b>	state-of-the-art classification	Not much accurate it select one of the best predictor from frequency table.
<b>Naive Bayesian</b>	Easy to implement. Less training data. Binary and multiclass classification problems.	Very strong assumption. Data scarcity. Continuous features.
<b>Decision Tree</b>	Variable screening or feature selection. Data preparation is easy. Explicit all possible alternatives and traces each alternative.	Due to over fit, they are prone to sampling errors.
<b>Linear Discriminant Analysis</b>	One of the best algorithms for face recognition Fast and portable Good to use when beginning a project.	Old algorithm. Some algorithms are much better prediction then this.
<b>Logistic Regression</b>	Handle nonlinear effects prediction is done both numerical and categorical	Boolean values only. Not suitable for predict the value of a binary variable
<b>K Nearest Neighbors</b>	Robust to noisy training data. No Training phase Can handle complex models easily	Hard to apply for larger dimension problems. Which type of distance metric to use. High cost
<b>Support Vector Machine</b>	Can be used on larger dimension. Different kernel function for various	Features greater then samples. Probabilities are not directly estimated.

	decision functions	
<b>Feed-forward neural network</b>	Complex functions very easily solved. To model non-linear dependencies Easy to maintain.	Cannot be used for smaller data available. Not good for arithmetic's and precise calculations
<b>Feed-back neural network</b>	back propagation travel in both directions Dynamic system which changes continuously.	Slow and inefficient. Can get stuck at local minima.
<b>Convolutional neural network</b>	Error is very less when compare to pervious. Classification problem on object recognition.	More hidden layers Time and space

*Table-I* Advantages and Disadvantages of different classifications

### III.CONCLUSION

This paper contains various classification techniques used in machine learning algorithms. A Classification is a method of predicting similar information from categorical or numerical datasets. Now a day's machine learning algorithms are became more popular for classification problems .This paper gives an introduction to most of the popular machine learning algorithms used for classification of pattern recognition.

### REFERENCES

- [1] W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0
- [2] M. Welling, "A First Encounter with Machine Learning"
- [3] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118-96174-2.
- [4] Chitra Nasa, Suman "Evaluation of Different Classification Techniques for WEB Data " International Journal of Computer Applications (0975 – 8887) Volume 52– No.9, August 2012.
- [5] Sandhya N. dhage, Sandhya N. dhage "A review on Machine Learning Techniques" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 4 Issue: 3
- [6] Ayon Dey "Machine Learning Algorithms: A Review" International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179.
- [7] S. B. Kotsiantis "Supervised Machine Learning: A Review of Classification Techniques" Informatica **31** (2007) 249-268

- [8] V. Sharma, S. Rai, A. Dev, “A Comprehensive Study of Artificial Neural Networks”, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN 2277128X, Volume 2, Issue 10, October 2012.
- [9] <https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python>