International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 4, Issue 5, May 2015 ISSN 2319 - 4847

A Survey on Machine Learning Classification Techniques

Nikhil Mandape

Department of Computer Engineering and Information Technology, College of Engineering, Pune, India

ABSTRACT

In the biomedical engineering, the assessment of the biological variations happening into the body of human is a challenge. Specifically, identifying the abnormal behavior of the human eye is really challenging due to the several complications in the process. The important part of the human eye is Retina, which can replicate the abnormal variations in the eye. Due to the requirement for disease identification techniques, the analysis of retinal image has gained sufficient significance in the research field. Since the diseases affect the human eye gradually, the identification of abnormal behavior using these techniques is really complex. These techniques are mostly reliant on manual intervention. But the success rate is quite low, since human observation is prone to error. These techniques must be highly accurate, since the treatment procedure varies for various abnormalities. Less accuracy may lead to fatal results due to wrong treatment. Hence, there should be such a automation technique, which give us high accuracy for disease identification applications of retina. This survey shows the study of different classification methods existed and their limitations which includes K-NN, SVM, Decision Trees, Naïve Bayes classifiers etc. But when we combine them to make an ensemble then classification accuracy can be improved.

Keywords: Machine Learning, Ensemble Technique, Datasets, Confusion Matrix, Accuracy.

1. INTRODUCTION

Since last two decades machine learning has become one of the mainstays of information technology. In machine learning, dataset (also called as sample) plays a crucial role and learning algorithm is used to discover and acquire knowledge from data. The learning and prediction performance will be affected by the quality or quantity of the dataset.

1.1 Dataset

A dataset comprises feature vectors (instances), where each instance is a set of features (attributes) describing an object. Let us look at example of Synthetic three-Gaussian dataset as shown in figure 1.1:

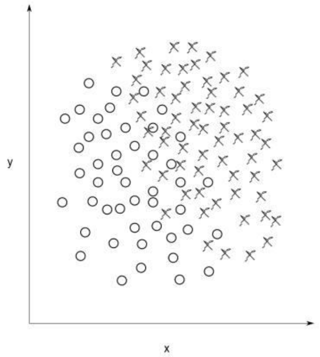


Figure 1 Synthetic three-Gaussian dataset

In this figure, the features x-coordinate, y-coordinate and shape describes one data point for each object, for example (0.8, 0.4, cross) is one feature vector. The quantity of features of the dataset is known as Dimension, for example in above sample the dimension is 3.

1.2 Training Set and Test Set

In machine learning, an unknown universal dataset is assumed to exist, which contains all the possible data pairs as well as their probability distribution of appearance in the real world. While in real applications, what we observed is only a subset of the universal dataset due to the lack of memory or some other unavoidable reasons. This acquired dataset is called the training set (training data) and used to learn the properties and knowledge of the universal dataset. In general, vectors in the training set are assumed independently and identically sampled from the universal dataset.

In machine learning, what we desire is that these learned properties can not only explain the training set, but also be used to predict unseen samples or future events. In order to examine the performance of learning, another dataset may be reserved for testing, called the test set or test data. For example, before final exams, the teacher may give students several questions for practice (training set), and the way he judges the performances of students is to examine them with another problem set (test set).

1.3 Validation set

Before testing, a learned model often needs to be configured, e.g., tuning the parameters and this process also involves the use of data with known ground-truth labels to evaluate the learning performance; this is called validation and the data is validation data. Generally, the test data should not overlap with the training and validation data; otherwise the estimated performance can be over-optimistic.

1.4 Model

One of the important tasks in machine learning is to construct good models from dataset. Generally a model is a predictive model or a model of the structure of the data that we want to build or determine from the data set, such as a support vector machine, a logistic regression, neural network, a decision tree etc. This method of creating models from data is called learning or training, which is accomplished by a learning algorithm. The learned model can be called a hypothesis (learner).

1.5 Learning Algorithms

There are different learning methods [10], among which the most common ones are supervised learning and unsupervised learning. In supervised learning, the goal is to predict the value of a target feature on unseen instances, and the learned model is also called a predictor. For example, if we want to predict the shape of the three-Gaussians data points, we call ‘cross’ and ‘circle’ labels, and the predictor should be able to predict the label of an instance for which the label information is unknown, e.g., (.4, .6). If the label is categorical, such as shape, the task is also called classification and the learner is also called classifier; if the label is numerical, such as x-coordinate, the task is also called regression and the learner is also called fitted regression model.

For both cases, the training process is conducted on data sets containing label information, and an instance with known label is also called an example. In binary classification, generally we use “positive” and “negative” to denote the two class labels. Unsupervised learning does not rely on label information, the goal of which is to discover some inherent distribution information in the data. A typical task is clustering, aiming to discover the cluster structure of data points. But the question is that “Can only single classifier give the best performance or shall we combine some techniques for better performance than single classifiers?” Now we will discuss about ensemble methods.

1.6 Ensemble methods

Ensemble methods[6] have become a major learning paradigm since the 1990s, with great promotion by two pieces of pioneering work. One is empirical [Hansen and Salamon,[5] 1990], in which it was observed that the best single classifier gives less accurate predications than the predictions by combination of a set of classifiers.

A simplified illustration is shown in Figure 1.2. The other is theoretical [Schapire, 1990], in which it was proved that weak learners are able to be enhanced to strong learners [9]. Since strong learners are desirable yet difficult to get, while weak learners are easy to obtain in real practice, this result opens a promising direction of generating strong learners by ensemble methods.

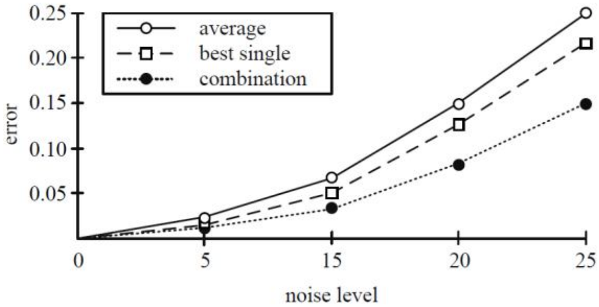


Figure 2 A simplified illustration of Hansen and Salamon [1990]’s observation: Ensemble is often better than the best single.

Generally, an ensemble [10] is constructed in two steps, i.e., generating the base learners, and then combining them. To get a good ensemble, it is generally believed that the accuracy of base learners should be more higher and diverse. It is worth mentioning that generally, the computational cost of constructing an ensemble is not much larger than creating a single learner. This is because when we want to use a single learner, we usually need to generate multiple versions of the learner for model selection or parameter tuning; this is comparable to generating base learners in ensembles, while the computational cost for combining base learners is often small since most combination strategies are simple.

2. LITERATURE SURVEY

While doing literature survey we found some existing classifier techniques used to examine performance of machine learning algorithms. Some limitations of these techniques are recognized.

2.1 K-Nearest Neighbor

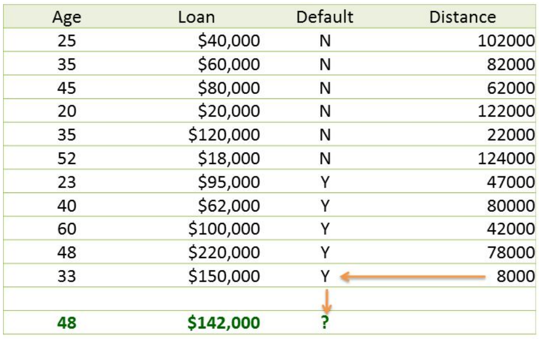
The k-Nearest Neighbors algorithm is a non-parametric method used for classification and regression in pattern recognition,. In both cases, the k closest training examples are there in the feature matrix. Whether k-NN is used for classification or regression, the output depends:

Class membership is the output in this algorithm. An object is classified by a majority vote of its neighbors, with the object being allocated to the class most mutual among its k nearest neighbors (k should be a small positive integer). When k = 1, then the object is simply allocated to the class of that single nearest neighbor.

The distance is calculated using one of the following measures:

2.1.1 Euclidean Distance

$$D=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

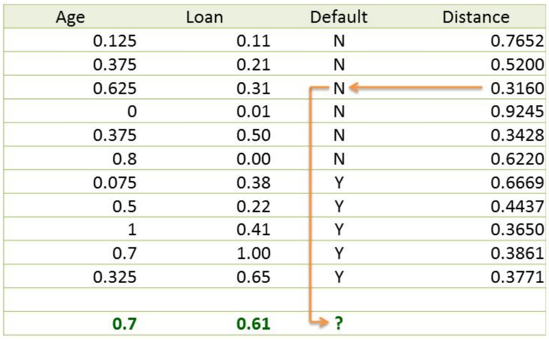


Age	Loan	Default	Distance
25	$40,000	N	102000
35	$60,000	N	82000
45	$80,000	N	62000
20	$20,000	N	122000
35	$120,000	N	22000
52	$18,000	N	124000
23	$95,000	Y	47000
40	$62,000	Y	80000
60	$100,000	Y	42000
48	$220,000	Y	78000
33	$150,000	Y	8000
48	$142,000	?	

Fig: Euclidian Distance

2.1.2 Standardized Distance

$$X_s = \frac{X - Min}{Max - Min}$$



Age	Loan	Default	Distance
0.125	0.11	N	0.7652
0.375	0.21	N	0.5200
0.625	0.31	N	0.3160
0	0.01	N	0.9245
0.375	0.50	N	0.3428
0.8	0.00	N	0.6220
0.075	0.38	Y	0.6669
0.5	0.22	Y	0.4437
1	0.41	Y	0.3650
0.7	1.00	Y	0.3861
0.325	0.65	Y	0.3771
0.7	0.61	?	

Fig: Standardized Distance

Limitations

- Knn Classification is time consuming.
- It is having great calculation complexity.
- This algorithm is fully dependent on training set.
- There is no weight difference between each class.

2.2 Support vector machine

Support Vector Machine (SVM) is a little example of learning system in view of statistical learning hypothesis. It maps a data test into a high dimensional feature matrix and tries to locate an ideal hyper plane that minimizes the recognition error for the training information by utilizing the non-linear transformation function. The center thought of a SVM is to minimize exact risk and the upper limit of expected risk, while at the same time enhance the capacity of the learning machine to sum up and to keep away from local minima, viably tackling the issue of over learning. With its great speculation execution in solving non-linear issues, the SVM has been effectively utilized as a part of handwriting recognition, face recognition, Feature selection, Chinese text location and Sketch retrieval. Limitations [12]

- In this technique there is strong randomicity in dataset.
- The training subset size in SVM is on large scale.
- The ensemble classifier of SVM is having high complexity.
- An SVM classifier is unbalanced on a training set of small size.
- If the number of feature dimensions is much greater than the training set size, then over fitting happens.

2.3 Discriminant Analysis

Discriminant Analysis is the basic method of classification in machine learning. It uses the ‘Classify()’ function in matlab. Supervised classification is also called as discriminant analysis and unsupervised classification is having another name as cluster analysis. Sometimes, along these lines, clustering is distinguished from classification. It is observed that discriminant analysis is “discrete prediction”, whereas regression analysis is “continuous prediction”. So a neural network with multilayer perceptron is an example of discriminant analysis; but used for continuous mapping and it is non-linear regression.

Limitations

- This algorithm does not work for the predictor variables which are incompletely measured.
- It is not intended to address the problem of missing data.
- Before analysis, testing of data is required.
- LDA is affected by non-normality, unequal dataset, and/or frequently small sample sizes make classification results unbalanced under resampling or cross-validation.

2.4 Decision Trees

A decision tree (or D-tree) is a classifier stated as a recursive division of the instance space. Composition of nodes that form a rooted tree is a Decision tree, meaning it is a directed tree with a node called “root” that has zero incoming edges but can have outgoing edges. All other type of nodes has just one incoming edge. A node with outgoing edges is called an internal/test/intermediate node. Remaining nodes are named as leaves (or decision or terminal nodes). In a decision tree, every intermediate node separates the instance space into two or more sub-spaces according to some discrete function of the input feature matrix values. In the easiest and regular case, each test considers a single attribute, such that the instance space is split permitting to the values of attribute. For attributes containing numeric values, the condition denotes to a range.

How do classification trees work?

1. For building a model, training data is used.
2. Tree creator determines:

a. Which variable to divide at a node and the splitting value.
b. Decision that where to stop (making a terminal note)

Limitations:

- Unlike Diagonal splits, parallel splits make some problems harder to learn.
- When applied to the full data set, decision trees tend to overfit training data which can give poor/fatal results.
- It is sometimes inefficient when splitting is done perpendicular to feature matrix axes.
- Extremely difficult to predict outside the lower and upper limits of the response variable in the training data.

2.5 Naive Bayes

Naïve bayes classifier is a part of family of simple probabilistic classifiers based on Bayes’ theorem in machine learning. The theorem is based on strong (naive) independent assumptions between the features set. It is a simple classification technique in which, models that assign class labels to problem instances, signified as vectors of feature matrix values, where the class labels are drained from some finite set.

It is a family of algorithms based on a common principle but not a single algorithm for training such classifiers. Common principle is: “All naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.” For example, if a fruit is red, round, and about 3" in diameter, then it may be considered to be an apple. This classifier considers each of these features to contribute individually to the possibility that this fruit is an apple, regardless of any possible correlations between the color, roundness and diameter features.

Limitations:

- Unbiased Learning of Naïve Bayes Classifiers is Impractical.
- An unrealistic number of training examples is required to learn Bayes classifiers for better performance.
- While naive Bayes frequently fails to produce a good estimation for the accurate class likelihoods, this may not be a condition for many applications. For example, the correct MAP decision rule classification is made by the naive Bayes classifier so long as the correct class is more probable than any other class.

3. CONCLUSION

In this survey we have studied different classification methods existed and their limitations. According to literature survey no single method is capable of giving accuracy as expected. So it is better to use the specific ensemble technique at specific time. For example Random Forest is ensemble of Decision Trees can be used for classification of retina images. Application of different ensembles gives us different performance results based on Accuracy and Confusion Matrix. Using different approaches choose the technique which gives us maximum accuracy.

REFERENCES

[1] Breiman, L., Bagging Predictors, Machine Learning, 24(2), pp.123-140, 1996.

[2] Thomas G. Dietterich “An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization” in Kluwer Academic Publishers, 1999.

[3] Xueyi Wang “A New Model for Measuring the Accuracies of” in IEEE World Congress on Computational Intelligence, 2012.

[4] Nikunj C. Oza and KaPSOnTumer “Classifier Ensembles: Select Real-World applications” in Elsevier, 2007.

[5] L.K. Hansen and P. Salamon, “Neural Network Ensembles,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, pp. 993-1001, 1990.

[6] Dietterich, Thomas G. "Ensemble methods in machine learning." In Multiple classifier systems, pp. 1-15. Springer Berlin Heidelberg, 2000

[7] Gonzalo Martı´nez-Mun˜ oz, Daniel Herna´ ndez-Lobato, and Alberto Sua´ rez, “An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation” Computer Science Department, EscuelaPoltie´cnica Superior, C/ Francisco Toma´s y Valiente, 11, Universidad Auto´noma de Madrid, Cantoblanco, 28049, Spain.

[8] Clifton D. Sutton, “Classification and Regression Trees, Bagging, and Boosting”, Department of Applied and Engineering Statistics, MS4A7, George Mason University, 4400 university drive, Fairfax VA 22030-4444 USA.

[9] R.E. Schapire. The strenght of weak learnability. Machine Learning, 5(2):197-227, 1990.

[10] Zhi-Hua Zhou, “Ensemble Methods- Foundations and Algorithms”.

[11] Sarwesh et al., “A Review of Ensemble Technique for Improving Majority Voting for Classifier”,International Journal of Advanced Research in Computer Science and Software Engineering 3(1), January - 2013, pp. 177-180

[12] Zhou Tao, “A new Classification algorithm Based on Ensemble PSO SVM and Clustering analysis”, Lu Huiling School of Science Ningxia Medical University Ningxia Yinchuan,P.R.China 750004

AUTHOR

Nikhil Mandape received the B.E. degree in Computer Engineering from MAEER’S MIT College of Engineering Pune, in 2011. He is currently pursuing his M.Tech. Degree in Computer Engineering from College of Engineering, Pune.