Generalized Discussion over Classification Algorithm under Supervised Machine Learning Paradigm

Sheenam Goel
M.Tech Scholar
Department of CSE
Shobhit University Meerut

Mamta, PhD
Associate Professor
Department of CSE
Shobhit University Meerut

ABSTRACT
In this paper, Learning is an important parameter for developing machines that are intelligent as well as efficient. The studies of virtual environment with parameters that are encountered periodically during run time of algorithm are studied effectively under machine learning domains. Optimized decision making floors the base of pattern recognition as a subarea of machine learning. Being influenced from theories of genetic sciences, cognitive learning, the efficiency of algorithms developed in this area is effectively exploited. However ensuring the adaptability of machines to have artificial thinking & generate optimum results when applied to domain of computer vision, various techniques have been correlated by means of diverse paradigm approach. Estimation of efficiency of one algorithm over other suitably forecast the optimum though not the best solution in terms of minimized error rate when applied to a problem statement. Although machine learning involves automation, but it imbibes human guidance to generate effective results and provides generalization on system so that they perform well on data patterns hidden in a problem space. The paper focuses on classical discussion over different techniques to be applied on areas of machine learning like classification and regression, two important aspects of learning over binary and multiclass problems. However the applicability of statistical models have grewed up with the deficiencies of lacking reasoning capabilities, handling categorical data and missing values with the major drawback of skipping reasoning and generalizing ability. So the advent of learning algorithm have revolutionize the performance of system by imbibing artificial data with knowledge applied from experience i.e. training machines in order to generate correct results. Classification problems have been widespread in both binary and multiclass datasets. So having employs this supervised approach for appropriate handling of such kind of problems and determine the effectiveness of each with its shortcomings are generalized in the paper. The paper will be focused on explanatory techniques of classification their discussion domains of applications so that when they are applied on data set, they generate effective results.

Keywords
Learning, Recognition, Classification, Binary and Multiclass

1. INTRODUCTION
The study of various machine learning model generally attempts to minimize error over unknown data classification while considering data from different paradigms. Classification model usually draw conclusion from observed values. Under learning methodologies, supervised learning contributes its roots in classification paranoma while clustering sprang from unsupervised criterion of learning. When applied significantly machine learning solution attempts to forecast significant values to problems under consideration by extraction of hidden data information. With the widespread acceptance of machine learning phenomena, various approximation techniques have found their applications in areas of computer vision. Prediction and correct voting is critical task in imbalance data multiclass classification [1].In multiclass problems every data instance belongs to a set of previously defined labels. Machine learning is however the application of artificial intelligence where information is processed through algorithms to manipulate statistical data. However it uses classical techniques that are too stringent for ongoing processing in Big Data era, where complexity of data is high and multifaceted. It provides increasing level of automation in knowledge engineering process replacing much time consuming human activity with auto techniques that improve efficiency by discovering and exploring regularities in training data [2].However data requirements are generally high to ensure accuracy in prediction. Using machine learning effectively and successfully ports down to a combination of knowledge, awareness and ultimately taking a scientific approach to overall process[3].With data collection using modernized tools there comes challenges for analysis ,classification and interpretation. These challenges base their presence upon factors like noise, high dimensionality of feature space, sample sparsity and many more. As data independent classifier doesn't exist, an appropriate strategy is needed to visualize data, process it, and develop classifier to achieve our target. The above components should provide flexibility, accuracy and error free predictions. Various classification strategies consider data set as a single unit, in spite of multiple discriminating attributes and complexities involved. Certain models that base upon traditional methods remain the bottleneck of these types of problems. Classification techniques roots from the domains of data mining wherein we determine classes or labels to the data set on which classification is implied. The entire strategy is based on two basic concepts:

A) **Model Generation:** Where a set of class labels are determined. A set of tuples that are involved in construction of model are termed as training set. On training set further computations are performed to enhance the usage of model.
B) **Model Usage:** It implies applying the classification property. The known label of test sample is compared with classified results from the model [4].However the training and testing samples are apart.

Various examples can be sited under the problem of classification as to determine whether a particular customer

will be disburse a credit from bank or not depending upon parameters like his ability to pay back, past history.. These types of problems are labeled under binary class problems. Here we train the network on the basis of some labeled data samples and ensure that the network correctly output the result in spite of missing data in presence of disturbing factors to determine correct results that are close to optimum. The learning algorithm is fed with sample data samples normally derived. Sample comprising of variables like a1, a2….an called predicates corresponding to target variable say b. The learned memory is stored in data structure, we refer here as model specific to algorithm employed. Various constrains need to be followed as features must carry sufficient information to predict value of b for which the value of these predicates must be between {-1… +1} and all samples must be evenly distributed over target variables. The above approach determines optimum result and ensures accuracy in classification. Therefore choice of appropriate algorithm is chosen on the basis of various factors as discussed in this paper.

2. CONCEPT OF DATA MINING

It is potential extraction of valid, new and relevant information from data. It aims to forecast hidden patterns embedded in the data under analysis. However with the rapid growth in data generation, handling with care and efficiency becomes prime factors of concern. The process is iterative and various steps of relevance include preparation of data and understanding the application, modeling of appropriate network to process the data, evaluate the results and deploy the model for future work. However redundant and missing values should be dealt effectively while mining important patterns from data under consideration. Data miming has evolved from intersection of various diverse fields like machine learning, database, Statistics and Artificial intelligence [5].These mining algorithm stimulate extraction of relevant information.

3. MACHINE LEARNING PARADIGMS

The fig 1 illustrates various machine learning domains and their brief classification of techniques involved.

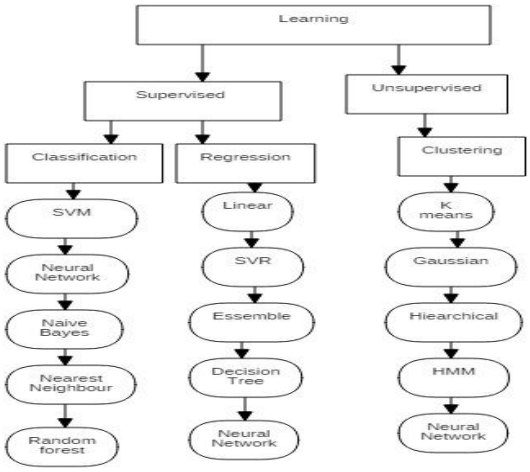

Fig1: Machine learning Paradigms

3.1 Insight of Supervised Machine Learning

Supervised Learning facilitates that both input and output are known in advance. Upon the availability of training data set, the supervised algorithm should be able to infer the response with all possible input parameters as shown in Fig 2. It occurs when the trainer provides classification for each example [6].To be able to solve a problem by employing supervised learning certain steps of relevance are followed:

1) Determine the various types of examples involved in training.
2) Accumulate training data set.
3) Determine a representation of learned function in terms of Input features.
4) Determine the corresponding algorithm used for learning.
5) Run the algorithm on training data set.
6) Evaluate the accuracy.

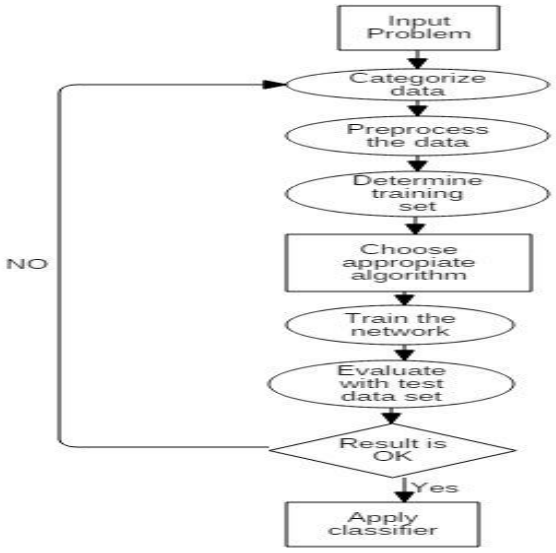

Fig 2: Insight of supervised machine learning

3.2 Depiction of Supervised Learning Mechanism

Illustrated in Fig 3 the mechanism involves known data set and Reponses that lead to the generation of appropriate learning model. Once the model is learned, the model is applied on testing data set to predict responses for unknown values of data in the problem presented to model.

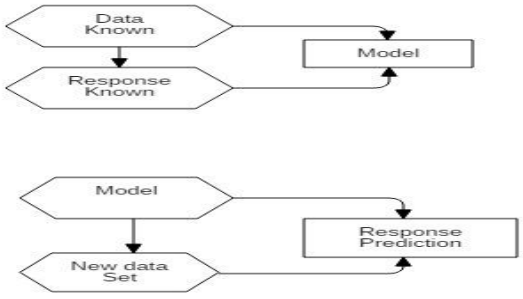

Fig 3: Mechanism of supervised learning model

4. CLASSIFICATION AS AN APPLICATION OF SUPERVISED LEARNING ALGORITHM

Classification deals with assigning class labels to the data set under consideration. As the amount of data is growing at a tremendous rate therefore these algorithms must be able to stimulate results effectively both when applied to structured and unstructured data Various algorithms like decision tree, neural network, nearest neighbor, SVM, Naïve bays etc can be employed for this purpose. Every technique employs a learning algorithm that determines as to which model fits best and establishes correspondence between attribute set and class label of input data. The model determined so far must be suitable enough to predict the class label estimate of records encountered for the first time. Therefore their key requirements are generalization ability.

4.1 Simplistic Approach of developing a classification model

The Fig 4 given below shows the general method of developing a model of classification wherein a model is developed on the suitable induction of training data set with learning algorithm. Once the model is generalized, this model is further applied on testing data set wherein classes are unknown and on the basis of past analysis and learning new deductions are facilitated.

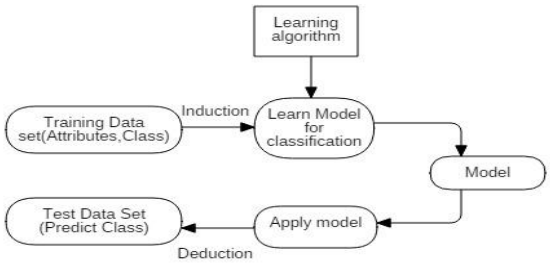

Fig 4: General classification model

4.2 Accuracy Test of Classification Model

The performance of model is dependent on the count of test records predicted by model either correctly or incorrectly. These predictions are illustrated by means of confusion matrix. A sample confusion matrix for a binary classification problem is shown in Table 1.

Table 1: Confusion Matrix

Class	Predicted Class	
	1	0
Class 1	a_{11}	a_{10}
Class 0	a_{01}	a_{00}

Where a_{ij} is a record from class I predicted to be of class j. For instance in the above example a11 denotes correct labeling from class 1 to class 1.A total of correct predictions are ($a_{11}+a_{00}$) and incorrect predictions are ($a_{01}+a_{10}$).However more accurate parameters of efficiency employ:

Accuracy = no of predictions that are accurate / Total no of predictions

$$=a_{11}+a_{00}/a_{11}+a_{00}+a_{01}+a_{10}$$

Error Rate = no of wrong predictions / Total no of predictions

$$=a_{10}+a_{01}/a_{11}+a_{00}+a_{10}+a_{01}$$

To ensure accurate results error rate must be minimized.

4.3 Classification Domains

Classification generally predicts the categorical class labels of new data sets under consideration based on past observations .The class labels may be binary in nature is in ex of spam detection on email, the answer to it is either yes or no. But the real world problems are not always binary. The predictive model that was learned during supervised learning algorithm implementation may assign a class value to an unlabeled instance presented in the training data set. However in case of character recognition scheme if the machine is able to recognize different letters in an handwriting specific to an individual with certain accuracy then on the other instance if different handwriting is presented to machine, it shall predict with same accuracy. To illustrate the concept of binary classification as shown in the Fig 5 a, if there are in total 10 samples where 4 are labeled with +ve class and remaining with negative class, they can be easily depicted on a 2D plane with labels Y1, Y2. On the other hand if classification predicts categorical labels then regression analysis facilitates relation between predicator and target variables to determine the outcome.eg: We want to predict the scores of Chemistry test of our students. If there exists a relationship between times spend in studying and test scores, we can use this correlation to predict test scores of students opting same subject in future. The figure 5 b illustrates the concept of linear regression .We use straight line that minimizes distance between sample points and fitted lines. Using slope learned from this data we generalize outcome of future data sets.

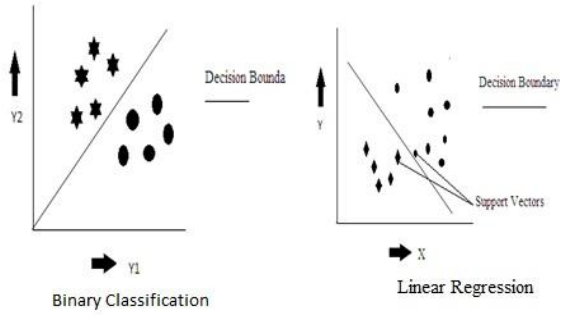

Fig 5a, 5b: Binary Classification and Linear Regression

5. MODELS OF CLASSIFICATION UNDER DISCUSSION

Various models are illustrated for classification with their pros and cons:

5.1. Decision Tree
5.2. Random Forest
5.3. Support Vector Machine
5.4. Neural network
5.5. Logistic Regression

5.1 Decision Tree

It is a widely employed data mining algorithm to facilitate classification i.e. assigning class labels for target data sets under consideration. This method both complements and supplements the old statistical analysis technique, various tools like neural network of data mining etc. Since it can be employed in both discrete and continuous variables these methods have found great importance in areas of data visualization and interpretations.

As it is non parametric in nature so while dealing with large sample size we usually study data as being partitioned into training and testing data sets. The training data set is usually employed to train the model and then the validity of model is evaluated against the test cases selected randomly to ensure the accuracy of model. The resulting leaves are given class label as a representative of target values. A sample decision tree is shown in Fig 6.1.

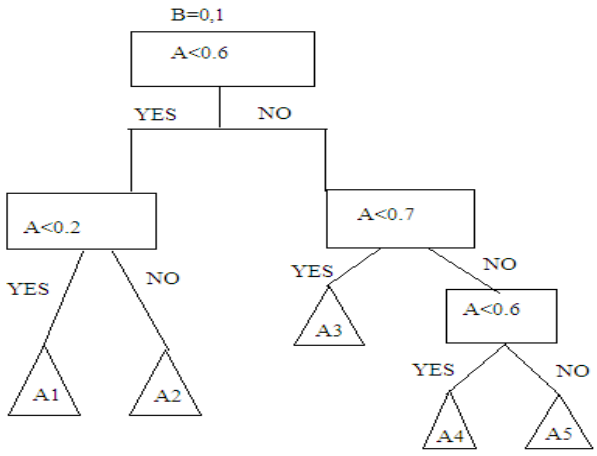

Fig 6.1: Decision Tree Illustration

The following decision tree is predicted upon target variables 0 and 1.The top node is the root node .The other nodes are internal nodes that represents possible choice made at the root whereas at lowest level are leaves that represent final decision after traversing through different paths in the tree as in eg. X1, X2 in fig 6.1. The paths from root to each internal node are represented using decision rule (if then).The input variables are utilized as splitting factor to divide records at root nodes or at internal nodes into two or more categories. The choice of input variables facilitate efficiency in determination of resulting child nodes after splitting depending on characteristics like gini index, entropy, classification index, twinning criterion. Etc. However complexity and robustness go hand in hand in selecting features while building statistical model. Having a complex model will result in being it less reliable to predict future records and it lacks generalisability. Therefore appropriate stooping rules are employed to avoid complexity in developing the model. However the model favors:

a) Handling of numerical and nominal attributes.
b) Dealing with missing values.
c) Dealing with data sets having multiple missing error values.

However beside the above factors this model works less efficiently when attribute interaction is highly complex designing of model tends to turn complex. Decision tree therefore aims to identify the best model for siding all records into different segments [7].

5.2 Random Forest

Random forest algorithm finds its applications in both classifications as well as regression domains. With the available training data sets having features, decision tree algorithm specified a set of rules that facilitate prediction on it. However to overcome the disadvantages of decision tree, random forest was proposed to handle missing, categorical values and avoid problems of over fitting in a better manner. However the random forest bases it prediction on two fold approach i.e. creation of random forest pseudo code and second to use it in facilitating prediction.

The pseudo code works as follows:

1. Selection of X features from a total of Y "features such that X<<Y
2. Amongst the selected X features select the root node M using best split point.
3. Split M in successive daughter nodes.
4. Repeat step 1 to 3 until single node remains.
5. Repeat step 1 to 4 until n no of trees is created.

To perform prediction we generally obtain test features and use the rules of each tree created above to predict the outcome. Since after the votes are calculated for each prediction and the highest voted targets are utilized for prediction. It is quite fast and scalable over range of applications. But it over fits the class [3].

5.3 Support Vector Machine

Another supervised machine learning algorithm that is widely employed in problems of classification and regression. Classification with it is centric on finding a hyper plane while plotting data items as points on n dimensional feature space. The hyper plane divides the available data set into separate classes as shown in the Fig 6.2:

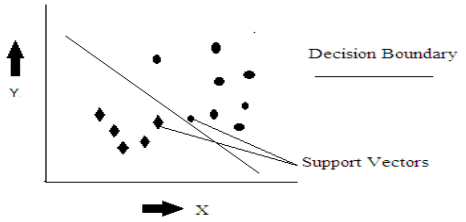

Fig 6.2: Concept of hyper plane

The data points that lie close to the hyper plane are support vectors .The aim of classifier is to separate the data points on the plane and ensure that the hyper plane so chosen has the greatest possible margin (i.e. the distance between nearest data points and hyper plane) so that the newly arrived data points are classified in the testing data set.

Various choices of hyper plane are proposed below in the fig 7(A, B, C):

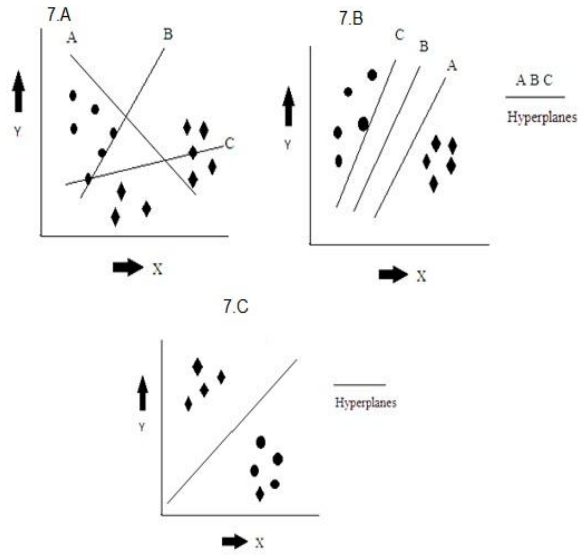

Fig 7.A shows that amongst the three hyper planes the one having maximum data points in its eccentricity is selected.

Fig 7.B shows that hyper plane selection is based on one having highest margin.

Fig 7.C illustrates how two class of data points having outliers, choosing one with maximum margin.

5.4 Neural Network

It is widely accepted classification technique which employs the artificial neural network that assembles electronic neurons in network in a network structure and process data records one at a time and enhance their learning process by comparing their record classification done arbitrary with actual classification of data set. The errors are propagated back to the network to ensure accuracy during future iterations. During training phase correct output are known for each record and each newly generated output is compared with actual values. The network trains itself by adjusting weights to predict correct labels for input. They are highly tolerant to noisy data and classify patterns for which network is not been trained. It has been advanced ion the form of ensemble methods, Ada boost and bagging .Performance advantage over single NN is achieved by these methods.

5.5 Logistic Regression

It tends to approximate probability (P $(y_q/s_p x_q)$) i.e. corresponding output based on certain input x_q where y_q is predicted output. A memory based model that is capable of self tuning on the basis of noise level of training data .Here the dependent variable is categorical in nature either being a binary outcome i.e. 0 ,1 or having multiple outcomes which are analyzed using multinomial logistic regression. This method is quite effective simple capable of extrapolating, competitive and has accurate results over neural network classifier. Global logistic regression can be extended to multiple categorical output values. However various scenarios of logistic regression are discussed in fig 8 given below:

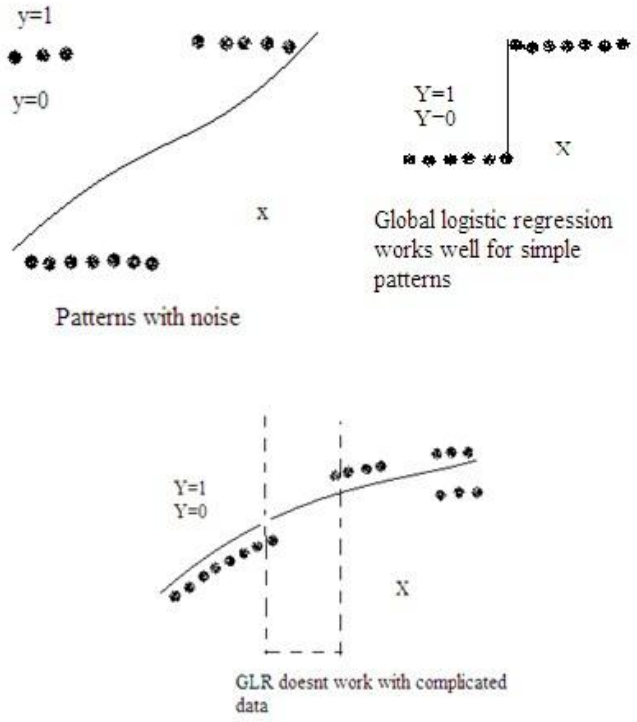

Fig 8 Scenario's of logistic regression

The input is represented on x axis and corresponding Boolean output on y axis either 0 or 1.The cross represent data points in memory. The above Fig 8 illustrates that global logistic regression works well with simple &noisy patterns but fails with complex patterns. To overcome the problem we either use feed forward multilayer preceptor neural network or localization paradigm.

6. CONCLUSION

In this paper we have reviewed over variety of classification algorithms with their pros and cons. When applied on variety of data sets different algorithms give different results each having accuracy improved factor over the other. The future work shall be based upon implementation and comparison of these algorithms on a problem statement that can be implied from real world. With these algorithms it is feasible to study under different circumstances which algorithm will excel in terms of efficiency and accuracy .So with less computation and time involvement machine will produce better results.

7. REFERENCES

[1] M Sahare , H Gupta , "A Review of Multi-Class Classification for Imbalanced Data", International Journal of Advanced Computer Research ,ISSN 2249-7277 ,Vol. 2,2012.

[2] Y Singh,PK Bhatia," A Review of Studies on Machine learning techniques",*IJCSS*,Vol.1.

[3] B Wujek,P Hall,F Güneş ," Best Practices for Machine Learning Applications", Paper SAS2360-2016.

[4] M Gupta,N Aggarwal," Classification Techniques Analysis", National Conference on Computational Instrumentation CSIO Chandigarh,March 2010.

[5] M Allahyari,S Pouriyeh, "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques", Xiv:1707.02919v2 [cs.CL] , Jul 2017.

[6] A Talwar,Y Kumar," Machine Learning: An artificial intelligence methodology", .International Journal Of Engineering And Computer Science ,ISSN:2319-7242 ,Vol. 2 ,2013.

[7] Y-y-Song,Y LU ," Decision tree methods: applications for classification & prediction", Shanghai Arch Psychiatry,2015.

[8] A Gupta,S Joshi," Study of Classification Algorithms used in Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 5 , 2014.

[9] Shotaro Matsumoto, Hiroya Takamura, and Manabu Okumura "Sentiment Classification Using Word Sub-sequences and Dependency Sub-trees" Advances in Knowledge Discovery and Data Mining ,301-311, 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam, May 18-20, 2005. Proceedings.

[10] T.-N. Do and F. Poulet, ``Random local SVMs for classifying largedatasets," in Future Data and Security Engineering SE-1, vol. 9446. Cham, Switzerland: Springer, 2015, pp. 3_15.

[11] B.Waske and J. A. Benediktsson, ``Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858_3866, Dec. 2007

[12]] K.-R.M¨uller, S. Mika, G. R¨atsch, K. Tsuda, and B. Sch¨olkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001

[13]] P. Laskov, C. Sch¨afer, and I. Kotenko. Intrusion detection in unlabeled data with quarter-sphere support vector machines. In *Proc. DIMVA*, pages 71–82, 2004.

[14] Kamal Nigam, John Lafferty ,Andrew McCallum," Using Maximum Entropy for Text Classification" IJCAI-99 Workshop on Machine Learning for Information Filtering, 1999, Pages 61-67 – Max Entropy

[15] L. Torrey and J. Shavlik, Handbook of Research on Machine Learning *Applications and Trends*. Hershey, PA: IGI Global, 2010.

[16] S. Haykin. Neural Networks: A comprehensive foundation, 2nd Ed. Prentice-Hall, 1999. D.P. Helmbold, D.D.E. Long, T.L. Sconyers, and B. Sherrod. Adaptive disk spin-down for mobile computers. *Mobile Networks and Applications*, 5(4):285–297, 2000.

[17]] S. Mendelson and A. Smola, editors. *Advanced Lectures on Machine Learning*, volume 2600 of *LNAI*. Springer, 2003.

[18] W. Zang, P. Zhang, C. Zhou, and L. Guo, ``Comparative study between incremental and ensemble learning on data streams: Case study," *J. Big Data*, vol. 1, no. 1, p. 5, 2014.

[19] M. K Warmuth, J. Liao, G. R¨atsch, Mathieson. M., S. Putta, and C. Lemmem. Support Vector Machines for active learning in the drug discovery process. *Journal of Chemical Information Sciences*, 43(2):667–673, 2003Decision tree methods: applications for classification and prediction

[20] [20] V.N. Vapnik. The nature of statistical learning theory. Springer Verlag, New York, 1995. U. von Luxburg, O. Bousquet, and G. R¨atsch, editors. Advanced Lectures on Machine Learning, volume 3176 of LNAI. Springer, 2004.