International Journal of Engineering & Technology, 7 (2.19) (2018) 31-35

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Study on Machine Learning: Elements, Characteristics and Algorithms

K Chokkanathan[1]*, S Koteeswaran [2]

Research Scholar,
Department of Computer Science and Engineering, School of Computing,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Avadi, Chennai-62, TamilNadu, India.
and
Assistant Professor,
Department of Computer Applications,
Madanapalle Institute of Technology & Science,
Madanapalle - 517 325, Andhra Pradesh, India.
[2] Associate Professor,
Department of Computer Science and Engineering, School of Computing,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Avadi, Chennai-62, TamilNadu, India.
**Corresponding author E-mail: kchokkanathan@outlook.com*

Abstract

Machine learning algorithms are used immensely for performing most important computational tasks with the help of sample data sets. Most of the cases Machine learning algorithms will provide best solution where the programming languages failed to produce viable and economically cost-effective results. Huge volume of deterministic problems are addressed and tackled by using the available sample data sets. Because of this now a days machine learning concepts are extensively used in computer science and many other fields. But still we need to explore more to implement machine learning in a specific field such as network analysis, stock trading, spam filters, traffic analysis, real-time and non-real time traffic etc., which may not be available in text books. Here I would like to discourse some of the key points that the machine learning researchers and practitioners can make use of them. These include shortcomings and concerns also.

Keyword: *Deterministic,network analysis, real time and non-real time traffic, spam filters.*

1. Introduction

From a give set of test information machine learning calculations or techniques will take in the projects consequently. Because of this adaptability the machine learning calculations were utilized as a part of software engineering and numerous different fields. Manual communication will be diminished however much as could reasonably be expected. Since it is gaining from the informational index consequently, it's connected in numerous applications, for example, misrepresentation identification, tranquilize configuration, stock exchanging, spam channels, web seek and numerous different fields. Late reports are demonstrating that these calculations are utilized as a part of an inventive way for information mining and information examination regions.These algorithms are the driving forces of next innovative generation [1]. There are diverse kinds of machine learning calculations exist, however I might want to center for the most part around most settled and broadly utilized Classification approach. Fundamentally classifier is a framework that takes a vector of discrete or persistent component esteems as information and yield as a solitary discrete esteem, the class. For instance parcel grouping characterizes ongoing and non-continuous bundles amid information transmission over the system.

2. Representing an Algorithm

Assume for a specific situation we may surmise that machine learning approach is great however the precarious is the means by which to pick one from the puzzling assortment of learning calculations. Fundamentally we ought to comprehend one thing that any machine learning calculation is the blend of three segments. Those parts are: portrayal, assessment and advancement.

Representation: whatever the classifier we are picking ought to be inside the premises of specific space and furthermore we require some formal dialect to speak to the classifier, with the goal that the PC can deal with productively. In the event that the classifier isn't in the speculation space it can't be educated.

Evaluation: There might be number of classifiers for a specific area. It's important to recognize and recognize great and appropriate classifier through the assessment procedure. There are two distinctive assessment methods like inward and outside assessment, both are important to get the upgraded arrangement in a characterization procedure.

Optimization: Among the classifiers we have to get the most astounding score for better conclusion. The best strategy is the decision of improvement system, which will give the effective yield to the student. At times there may in excess of one ideal. It's

the common obligation regarding students to pick the standard enhancers, which are later traded with hand crafted ones.
Following Table-1 displays the examples commonly representing these three components[2]. In k-nearest neighbor classifier finds the k most comparative preparing informational collections and foreseeing the standard class among them. Hyper plane technique predicts the class with the most elevated esteemed combination.Decision tree will test every single hub, branch and will do class expectations at the leaves and so on.

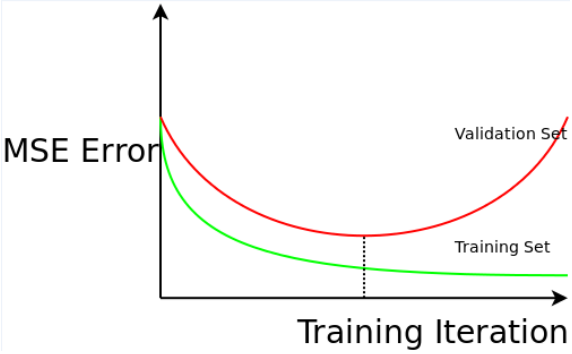
Table 1: The three components of learning algorithms.

Representation	Evaluation	Optimization
Instances K-nearest neighbor Support vector machines Hyper planes Naïve Bayes Logistic regression Decision trees Set of rules Propositional rules Logic programs Neural networks Graphical models Bayesian networks Conditional random fields	Accuracy / Error rate Precision and recall Squared error Likelihood Posterior probability Information gain K-L divergence Cost / Utility Margin	Combinational optimization Greedy search Beam search Brach-and-bound Continuous optimization Unconstrained Gradient descent Conjugate gradient Quasi-Newton methods Constrained Linear programming Quadratic programming

3. Generalization in Machine learning

Internet of Things genuinely is a system of real world applications with real time interactions. Machine to Machine (M2M) communications is the improvement in the initial level of IoT which is having unique traits with the deployment contexts and contribution [9]. The best networking is the Wireless Area Network (WAN) which works without human intervention [12][13]. The varieties of threats within the security of IoT are shown in the figure 2.
Generalization a vital trademark for machine learning arrangements and assuming fundamental part in getting ideal arrangements, so we can't stick just with preparing set past that we need to think about information in a more extensive territory. Regardless of that how much information we have for preparing yet additionally we need to see those correct informational indexes again at test time. On the off chance that we have 100 words we have to create 2^{100} conceivable diverse data sources. We can't depend aimlessly on preparing informational collection in light of the fact that more often than not will make a few deceptions. Another critical factor is picking the classifier superior to anticipating. So regardless of whether we depend some body to fabricate a classifier, guarantee that a portion of the information to our self and test the classifier they give you on it. When we utilize distinctive informational index and amid the tuning procedure there are number of conceivable outcomes of sullying our classifier. This can be mitigated by doing cross-approval: haphazardly isolating your preparation information into (say) ten subgroups, holding out every one while investigating the rest, attempting each refined classifier on the examples it didn't see, and averaging the results to perceive how well the particular parameter setting does.
At the underlying stage, it was not acknowledged to keep the preparation and test information independently. Since in those days we got extremely constrained measure of informational index portrayals so the mistakes may not be vast and identifiable. Presently a days with a portion of the adaptable classifiers like choice tree and direct classifiers it is required to isolate the informational collection entirely. Generalization is real goal for machine learning in light of the fact that the results are supporting to deliver the ideal arrangements. From streamlining concern, it isn't obligatory to enhance completely toward the start, rather a neighborhood improvement will be superior to the worldwide advancement.

The accompanying diagram (Graph-1) is an extremely normal and surely understood to all information researchers. MSE approval diagram demonstrates that if the MSE of your preparation set goes too low, there is where the MSE of your cross approval set will build, which means poor anticipating power. This is where the calculation winds up finished fit and it loses the ability to sum up. This is the place machine learning turns out to be extremely nuanced and requires the understanding and experience of an information researcher [3].



Graph-1: Only Data will not be supportive

As a result of Generalization, whatever it might be the volume, information alone won't be sufficient to find out about the framework. For a specific area we may know lakhs of information yet at the same time we have to know more to discover that space adequately. Without additional data, there is only no real way to do this that beats flipping a coin. This perception was first made (in fairly unique shape) by the thinker David Hume more than 200 years prior, however even today numerous mix-ups in machine taking in originate from neglecting to welcome it. Each student must epitomize some information or suppositions past the information it's given keeping in mind the end goal to sum up past it. This was formalized by Wolpert in his renowned "no free lunch" hypotheses, as per which no student can beat arbitrary speculating over every conceivable capacity to be educated [4].
It's more hopeless news that how to master something utilizing a classifier. Luckily this present reality capacities are not drawn consistently from the arrangement of all scientifically encouraging capacities. We can understand one thing that why this machine learning is best? The reason is smoothness, cases with comparable true classes, constrained conditions and restricted multifaceted nature. Another gainful thing is little measure of information will deliver enormous measure of yield. The space with which we are working, the portrayal what we are utilizing is most strong in communicating our insight. Generally occasion based techniques may a best decision for portrayal. For instance on the off chance that we have great introduction in likelihood at that point to speak to our area we can make utilization of graphical models which will be honorable fit. If we know the prerequisites then “IF...THEN’’ rules will be the best option. The most useful learners in such manner are those that don't simply have suppositions hard-wired into them, yet enable us to state them unequivocally, fluctuate them broadly, and join them naturally into the learning (e.g., using first-order logic [5] or grammars [6]).Machine isn't enchantment and it's not something from nothing. We can't get more from less. We need to fabricate everything starting with no outside help. Learning is more similar to cultivating, which gives nature a chance to do the greater part of the work. Agriculturists join seeds with supplements to develop crops. Learners join information with information to develop programs [2].

4. Error or Noise due to Complexity

In the event that we don't have adequate information about information and classifier will present arbitrary blunders in the information and learning framework. This issue is for the most part known as finished fitting. This over fit will come in numerous

structures that are not instantly observable. At the point when your student yields a classifier that is 100% exact on the preparation information however just half precise on test information, when in truth it could have yield one that is 75% precise on both, it has over fit. One approach to comprehend over fitting is by breaking down generalization mistake into inclination and change[7].Reliably to find out about the wrong thing by the students is Bias. Propensity to find out about arbitrary things regardless of the genuine flag is known as Variance.
Figure 1 delineates this by a relationship with tossing darts at a board[7]. A direct student has high predisposition, since when the outskirts between two classes isn't a hyper plane the student can't prompt it. Decision trees don't have this issue since they can speak to any Boolean capacity, however then again they can experience the ill effects of high difference: Decision trees learned on various preparing sets produced by a similar wonder are regularly altogether different, when in actuality they ought to be the same. Comparative thinking applies to the decision of streamlining technique: pillar look has bring down predisposition than ravenous inquiry, however higher difference, since it tries more hypotheses. Thus, contrary to intuition,a more powerful learner is not necessarily better than a lesspowerful one.

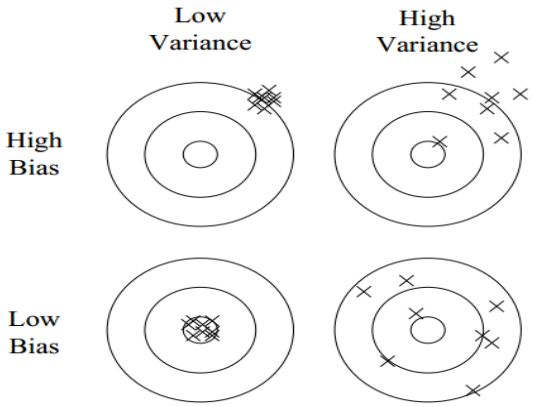


Fig 1: Bias and variance in dart-throwing.

For example by using it to choose the best size of decision tree to learn. But it's no panacea, since if we use it to make too many parameter choices it can itself start to over fit[8]. There is no single approach to solve the over fitting problem. However one of the approach is regularization term. That is penalize classifiers with more structure, thereby favoring smaller ones with less room to over fit. There is a wrong assumption that over fitting is caused by noise, but it's not write thing, even though there is no noise severe over fitting can occur.For example, assume we take in a Boolean classifier that is only the disjunction of the illustrations marked "valid" in the preparation set. (As it were, the classifier is a Boolean recipe in disjunctive typical shape, where each term is the conjunction of the component estimations of one particular preparing illustration) [2].

5. Perception Flops in High Magnitudes

Once the overfitting is fathomed, the following prompt issue is the scourge of dimensionality. At the point when there is low measurements the framework will react legitimately, if the info is high dimensional then the framework needs more recalcitrant assets. In machine taking in the response more in the event of high-measurements. In machine learning issues that include taking in a "condition of-nature" (perhaps an unending dispersion) from a limited number of information tests in a high-dimensional element space with each element having various conceivable qualities, a tremendous measure of preparing information are required to guarantee that there are a few examples with every blend of qualities. With a settled number of preparing tests[9], the prescient power decreases as the dimensionality increments, and this is known as the Hughes impact or Hughes marvel (named after Gordon F. Hughes)[10].
It is anything but difficult to build a classifier in at least two measurements. Through visual investigation we can locate a reasonable boondocks among various classes. In any case, in high measurements we can't comprehend what is occurring in the framework. With the goal that it's hard to propose and develop the great classifier. We may imagine that gathering more data will be strong however in actuality they give no extra data about the class. Their benefits are remunerated by the scourge of dimensionality. Luckily, there is an impact that halfway checks the revile, which may be known as the "gift of non-consistency." In many applications illustrations are not spread consistently all through the occasion space, but rather are focused on or almost a lower-dimensional complex. For instance, k-closest neighbor works great for written by hand digit acknowledgment despite the fact that pictures of digits have one measurement for each pixel, on the grounds that the space of digit pictures is considerably littler than the space of every single conceivable picture. Students can verifiably exploit this lower successful measurement, or calculations for unequivocally decreasing the dimensionality can be utilized.

6. Depending the Hypothetical Promises

The vast majority of the machine learning papers are promising hypothetically. Major of the sorts are predetermined on the quantity of models expected to guarantee great speculation. Enlistment is generally appeared differently in relation to reasoning: in derivation you can ensure that the conclusions are right; in acceptance what happens next is anyone's guess. Or on the other hand such was the tried and true way of thinking for a long time.
In the event that the hypothetical probabilities are summed up well then it is conceivable to ponder for the student. On the off chance that the student has all the more preparing informational collection then he can deliver high likelihood which restores a summed up hypothetical outcomes. To enhance the genuine classifier execution we require all the more preparing informational collection which will build the likelihood of student yields. On the off chance that we lessen the hypothetical space in preparing informational index naturally the conceivable outcomes that it incorporates the genuine classifier will be diminished.
At that point significant part of speculative space in machine learning is a wellspring of understanding and main thrust for calculation develop. In this perspective they vital and helpful. On account of these both hypothesis and useful ability machine learning progress more for the past and expected years.

7. Features Are Deciding Factors

A definitive inquiry a portion of the machine learning based undertakings are succeeded and some are fizzled what is the purpose for this preoccupation. The critical factor is highlights utilized as a part of every framework. In the event that we an ever increasing number of connected highlights with the class, learning is simple. In the event that mind boggling it's hard to learn. Rather than utilizing uncooked information we can develop highlights from those information later can be utilized for learning reason. For this we have to put more exertion. It joins some specialized stuffs like instinct, imagination are more essential. It's exceedingly tedious procedure to assemble information, coordinate it, clean it and preprocess it and we need to do numerous experimentation exercises for building a decent outline. It's an intelligent procedure of running the student, breaking down the outcomes, changing the information and repeating them. Highlight building is more troublesome on the grounds that it's area particular, while students can be to a great extent broadly useful. Nonetheless, there is no sharp boondocks between the two, and this is another reason the most helpful students are those that encourage consolidating knowledge[11].

8. Additional Data Leads to a Smart Algorithm

At the point when our classifier isn't delivering the exact yields as our desires, there might be two conceivable reasons: we need to go for better learning calculation, or else get more information. The vast majority of the students are focusing on first reason, yet the snappiest way to progress is regularly to simply go for more information. As a dependable guideline, an idiotic calculation with parts and heaps of information beats an astute one with humble measures of it.

The rationale behind the idea that more information beats better calculations is somewhat unpretentious. Say, for instance, we trust the connection between two factors –, for example, number of pages saw on a site and percent probability of a site guest to make a buy – is direct. Having more information focuses would enhance our gauge of the fundamental direct relationship.The following to graphs[12]in Figure-2, show that more data will give us a more accurate and confident estimation of the linear relationship.

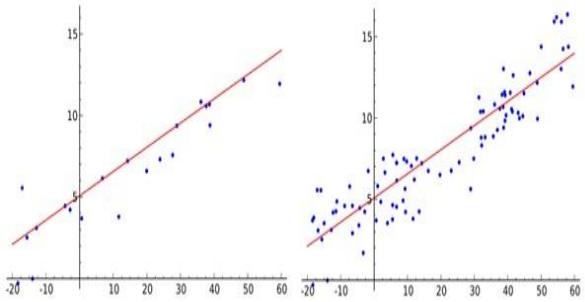


Figure-2 Estimating linear relationship

There are three imperative variables which are main thrust of machine figuring out how to an effective way. They are time, memory and preparing information. Prior days it was hard to get information, now we have a lot of information and issue is a great opportunity to process those information and how/where to suit them. Be that as it may it is limiting the procedure when we have gigantic volume of informational collection and furthermore enhancing the exactness. Some portion of the appropriate response is to concoct quick approaches to learn complex classifiers, and for sure there has been amazing advancement in this direction[12]. There is no sharp outskirts between planning students and learning classifiers; rather, any given bit of information could be encoded in the student or gained from information. So machine learning ventures frequently end up having a noteworthy segment of student plan, and specialists need some ability in it [13].

9. Absorb More Models Instead of One

There are different users having different opinion on different learning algorithms. They are assigning different priority for them based on their performance. A systematic comparison shows that learners behave differently from one application to another application. However the efforts are kept in finding the best one from many variations. But then researchers noticed that, if instead of selecting the best variation found, we combine many variations, the results are better—often much better—and at little extra effort for the user

Creating such model ensembles is now standard [14]. In the simplest technique, called bagging, we simply generate random variations of the training set by resampling, learn a classifier on each, and combine the results by voting. This works because it greatly reduces variance while only slightly increasing bias. In boosting, training examples have weights, and these are varied so that each new classifier focuses on the examples the previous ones tended to get wrong. In stacking, the outputs of individual classifiers become the inputs of a "higher-level" learner that figures out how best to combine them.

10. Conclusion

Like there are some vital fascinating factors about machine learning framework. Today the vast majority of the research works depend on machine learning calculations. Relies upon the space either a client can select managed or un supervised algorithms. There are such huge numbers of ongoing applications are utilizing this idea for classification reason. For instance, suppose that we need to prepare a ML model to foresee if an email is spam or not spam. Amid the information transmission we can group that whether the parcel data is real time or non-real-time. There are such a significant number of asset to find out about machine learning algorithm. Like any discipline, machine learning has a lot of "folk wisdom" that can be hard to come by, but is crucial for success. This article summarized some of the most salient items. A good place to learn more is my book The Master Algorithm, a non-technical introduction to machine learning15. There is also a treasure trove of machine learning lectures at http://www.videolectures.net. A widely used open source machine learning toolkit is Weka [16].

References

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers. Big data: The nextfrontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute, 2011.

[2] Pedro Domingos, "A Few Useful Things to Know about Machine Learning", Communications of the ACM, 2012.

[3] http://insidebigdata.com/2013/12/12/tech-tip-generalization-machine-learning.

[4] D. Wolpert. The lack of a priori distinctions between learning algorithms. Neural Computation,8:1341–1390, 1996.

[5] M. Richardson and P. Domingos. Markov logic networks. Machine Learning, 62:107–136, 2006.

[10] Classification for Remote Sensing". Mathematical Geosciences 40 (4): 409.

[11] Thomas Oommen, Debasmita Misra, Navin K. C. Twarakavi, Anupma Prakash, Bhaskar Sahoo, Sukumar Bandopadhyay, An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing, Volume 40, Issue 4, pp 409-424,May 2008.

[12] J. Tenenbaum, V. Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. Science, 290:2319–2323, 2000.

[13] G. Hulten and P. Domingos. Mining complex models from arbitrarily large databases in constant time. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 525–531, Edmonton, Canada, 2002. ACM Press.

[14] D. Kibler and P. Langley. Machine learning as an experimental science. In Proceedings of the Third European Working Session on Learning, London, UK, 1988. Pitman.

[15] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting and variants. Machine Learning, 36:105–142, 1999.

[16] R. Kohavi, R. Longbotham, D. Sommerfield, and R. Henne. Controlled experiments on the Web: Survey and practical guide. Data Mining and Knowledge Discovery, 18:140–181, 2009.

[17] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers. Big data: The next frontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute, 2011.

[6] W. Cohen. Grammatically biased learning: Learning logic programs using an explicit antecedentdescription language. Artificial Intelligence, 68:303–366, 1994.

[7] P. Domingos. A unified bias-variance decomposition and its applications. In Proceedings of the Seventeenth International Conference on Machine Learning, pages 231–238, Stanford, CA, 2000. Morgan Kaufmann.

[8] A. Y. Ng. Preventing "overfitting" of cross-validation data. In Proceedings of the Fourteenth International Conference on Machine Learning, pages 245–253, Nashville, TN, 1997. Morgan Kaufmann.

[9] Oommen, T. .; Misra, D. .; Twarakavi, N. K. C.; Prakash, A. .; Sahoo, B. .; Bandopadhyay, S. . (2008). "An Objective Analysis of Support Vector Machine Based