International Journal of Latest Trends in Engineering and Technology
Special Issue April-2018, pp. 016-023
e-ISSN:2278-621X

SURVEY ONDIFFERENT MACHINE LEARNING ALGORITHMS

S.A.Sarumathi[1], V.M.Sivagami[2]

Abstract---According to a recent study, a large number of techniques has been developed so far to tell the diversity of machine learning. Machine Learning (ML) has developed gradually by the few computer enthusiasts who exploits the possibility of computers learning to play games, data and a part of Mathematics (Statistics) that considered computational approaches, not only provided the necessary base for statistical-computational principles of learning procedures, but also has developed various algorithms that are regularly used for text interpretation, pattern recognition, sentiment classification, cyber security intrusion detection and a many other commercial purposes and has led to a separate research interest in data mining to identify hidden regularities or irregularities in social data that growing by second. Machine learning is categorized into supervised, unsupervised and reinforcement learning.
Keywords: Machine Learning, Regression, Techniques, Statistics, Data, Algorithms

1. INTRODUCTION

Machine learning is a paradigm that may refers to tell how to automatically find a good predictor based on past experiences (which in this case is previous data) to improve future performance[1]. The primary goal of Modern Machine Learning is highly accurate Predictions on test data [2]. Although you might argue that machine learning has been around as long as statistics has, it really only became a separate topic in the 1990's [5]. According to the nature of the available data, the two main categories of learning tasks are: supervised learning when both inputs and their desired outputs (labels) are known and the system learns to map inputs to outputs and unsupervised learning when desired outputs are not known and the system itself discovers the structure within the data. Examples of classification algorithms are k-nearest neighbour, logistic regression and Support Vector Machine (SVM) while regression examples include Support Vector Regression (SVR), linear regression, and polynomial regression. Some algorithms such as neural networks can be used for both, classification and regression [3]. There are several applications of Machine Learning; the main of all is Data mining.

2. DATA MINING

Data mining is moderately proving itself as an important tool for people who wish to analyze all this data for patterns. One of the most famous examples is from the late 1970s, when data mining proved itself as potentially important for both scientific and commercial purposes on a particular test application of diagnosing diseases in soybean plants [MC80] [5]. Data mining uses many machine learning methods, here are the algorithms:

1. C4.5
2. k-means
3. Support vector machines
4. Apriori
5. EM
6. Page Rank
7. AdaBoost
8. k-NN
9. Naive Bayes
10. CART

2.1 Most Common Data Mining Algorithm and Techniques

We have two different sections, each with a specific theme:

a. Classical Techniques: Statistics, Neighborhoods and Clustering
b. Next Generation Techniques: Trees, Networks and Rules [6]

2.2 Classical Techniques: Statistics, Neighborhoods and Clustering

1. The Classics

[1] PG.Scholar,Department of Information Technology , SriVenkateswara College of Engineering, Sriperumbudur, Tamilnadu India
[2] AssociateProfessor, Department of IT, SriVenkateswara College of Engineering, Sriperumbudur, Tamilnadu, India

This section should help the user to understand the rough differences in the techniques and at least enough information to be dangerous and well armed enough to not be confused by the vendors of different data mining tools.The main techniques that we may discuss here are the one that are used 99.9% of the time on current business problems [6].

2. Statistics

Machine learning and statistics are closely related fields. According to Michael I. Jordan, the ideas of machine learning, from methodological principles to theoretical tools, have had a long pre-history in statistics. He also suggested the term data science as a placeholder to call the overall field [28]. By strict definition "statistics" or statistical techniques are not data mining. However, statistical techniques are driven by the data and are used to discover patterns and build predictive models [6].

3. Clustering and Nearest Neighbor Prediction

K-means is a clustering algorithm that tries to partition a set of points into K sets (clusters) such that the points in each cluster tend to be near each other. It is unsupervised because the points have no external classification [26]. The Clustering are of two types: 1. Hierarchical and Non-Hierarchical Clustering [6]. There are two main types of hierarchical clustering algorithms:

Agglomerative - Agglomerative clustering techniques start with many clusters as where each cluster contains just one record and are merged together which are nearer to the next largest cluster. This merging is continued until a hierarchy of clusters is built with just a single cluster.

Divisive - Divisive clustering techniques take the opposite approach from agglomerative techniques. These techniques start with all the records in one cluster and then try to split that cluster into smaller pieces and then in turn to try to split those smaller pieces.

2.3 Next Generation Techniques: Trees, Networks and Rules

1. Decision Tree

A decision tree is a predictive model that, as its name implies, can be viewed as a tree. Specifically each branch of the tree is a classification question and the leaves of the tree are partitions of the dataset with their classification.You may notice some interesting things about the tree:

- It divides up the data on each branch point without losing any of the data.
- It is pretty easy to understand how the model is being built [5].

2. Neural Network

Neural networks are very powerful predictive modeling techniques which provide highly accurate predictive models that can be applied across a large number of different types of problems.There are two ways that these shortcomings in understanding the meaning of the neural network model have been successfully addressed:

- The neural network is package up with expert consulting services.
- This allows the neural network to be carefully crafted for one particular application and once it has been proven successful it can be used over and over again without requiring a deep understanding of how it works. [5].

3. CLASSIFICATION AND REGRESSION DATA MINING TECHNIQUES

A classification technique is a systematic approach to building classification models for training and testing data sets. The classification technique predicts the target class for each data point. For example, patients can be classified as “high risk” or “low risk” patient on the basis of their disease pattern using data classification approach. The main advantages and disadvantages of different classification techniques summary was given below as cited in [10]

Table 2.1 Advantages and disadvantages of different classification techniques

Methods	Advantage	Disadvantage
K-NN	1. It is easy to implement and faster	1. It requires large space. Sensitive to noise.
Decision Tree	1. There are no requirements of domain knowledge in the construction of decision tree. 2. It minimizes the ambiguity of complicated decisions and assigns exact values to outcomes of various actions.	1. It is restricted to one output attribute. 2. It generates categorical output. 3. If the type of dataset is numeric than it generates a complex decision tree.
Support Vector Machine	1.Better accuracy and Easy to handle complex nonlinear data points. 3. Over fitting problem is not as much as other methods.	1. Computationally expensive and consume much time. 2. The main problem is the selection of right kernel function. And shows different results. 4. SVM was designed to solve the problem of

		binary class.
Neural Network	1. Easily identify complex Relationships between dependent and independent variables.	1. Over-fitting. 3. The processing of ANN network is difficult to interpret and require high processing time if there are large neural networks.
Bayesian Belief Network	1. It makes computations process easier and Have better speed and accuracy.	1. It does not give accurate results in some cases.

4. MACHINE LEARNING ALGORITHMS

Machine learning, which is a field of research. It has found to be highly interdisciplinary field which acquires and constructs upon ideas from statistics, computer science (engineering), optimization theory, and numerous other disciplines of science and mathematics. Broadly, there are three main types of Machine Learning Algorithms:

4.1 Supervised Learning

This algorithm consists of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Machine learning algorithm makes predictions on given set of sample whereas Supervised learning algorithms searches for patterns within the value labels assigned to data points. The training process continues until the model achieves a desired level of accuracy on the training data. Examples-Linear Regression, Logistic Regression, Decision Trees, Naïve Bayes Classification [2][30].

4.2 Unsupervised Learning

In this algorithm, we do not have any target or outcome variable to predict/estimate [30]. Input data is not named and not have a known result. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention [1]. Unsupervised learning techniques are used for learn complex, highly non-linear models with millions parameters to used large amount of unlabeled data. Examples of Unsupervised Learning: Apriori algorithm, K-means, Fuzzy clustering, Hierarchical clustering [30].

4.3 Reinforcement Learning

Using this algorithm, the machine is trained to make specific decisions. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Reinforcement learning (sometimes called unsupervised learning) refers to a brand of learning situation where a machine should learn to behave in situations where feedback is not immediate.In this the machine is exposed to an environment where it trains. Example of Reinforcement Learning: Markov Decision Process [30].

5. SURVEY ON STATISTICAL METHOD ANALYSIS

- Non-Parametric
- Semi-Parametric
- Parametric

5.1 Non-Parametric Method

In non-parametric methods, an empirical estimate of the survival function is obtained using Kaplan-Meier (KM) method, Nelson-Aalen estimator (NA) or Life-Table(LT) method [7]. Nonparametric methods seek to best fit the training data in constructing the mapping function, whilst maintaining some ability to generalize to unseen data. As such, they are able to fit a large number of functional forms.Some more examples of popular nonparametric machine learning algorithms are [27]:

- k-Nearest Neighbors
- Decision Trees like CART and C4.5
- Support Vector Machines It also consists of some specific methods such as [7]:
- Kaplan-Meier
- Nelson-Aalen
- Life-Table

1. Kaplan-Meier

The Kaplan–Meier estimator, also known as the product limit estimator, is a non-parametric statistic used to estimate the survival function from lifetime data. The Kaplan–Meier estimator is one of the most frequently used methods of survival analysis. The Kaplan-Meier estimate is also called as “product limit estimate”.

Advantage

- The estimate may be useful to examinerecovery rates, the probability of death, and the effectiveness of treatment.

Disadvantage

- It is limited in its ability to estimate survival adjusted for covariates; parametric survival models and the Cox proportional hazards model may be useful to estimate covariate-adjusted survival [24].

2. Nelson-Aalen

The Nelson-Aalen estimator is a nonparametric estimator which may be used to estimate the cumulative hazard rate function from censored survival data or incomplete data. It is also used in theory of survival to estimate the cumulative number of expected events[9].

Advantages

- Nelson-Aalen estimator performs better when the sample size is small.
- The Nelson–Aalen estimator is always larger than log of the Kaplan– Meier estimator.

Disadvantages

- They are not very different, as long as the number of individuals at risk remains large
- They do not provide effective estimates and cannot generally be used to assess the effect of multiplefactors of interest (multivariable models) [11].

3. Life-Table

A life table presents the proportion of surviving, the cumulative hazard function, and the hazard rates of a large group of subjects followed over time. A life table is constructed from a set of grouped or ungrouped failure data. The life-table method competes with the Kaplan-Meier product-limit method as a technique for survival analysis.

Advantage:For large samples, the life-table method is popular in providing a simple summary of a large set of data [12].

Advantages of Non-Parametric

- Capable of fitting a large number of functional forms.
- Can result in higher performance models for prediction and more efficient.

Disadvantages

- Require a lot more training data to estimate the mapping function.Difficult to interpret yields inaccurate estimates
- More of a risk to overfit the training data and it is harder to explain why specific predictions are made [27].

5.2 Semi-Parametric Methods

Cox regression method is described as semi-parametric method since the distribution of the outcomeremains unknown even if it is based on a parametric regression model. They are often used in situations where the fully nonparametric model may not perform well or when the researcher wants to use a parametric model but the functional form with respect to a subset of the regressors or the density of the errors is not known [7]. Here, some of the specific methods are

- Cox model
- Regularized Cox
- Cox Boost
- Time-Dependent Cox

1. Cox Regression

Cox regression (or proportional hazards regression) is method for investigating the effect of several variables upon the time a specified event takes to happen. The method does not assume any particular "survival model" but it is not truly nonparametric because it does assume that the effects of the predictor variables upon survival are constant over time and are additive in one scale [13].

Advantages

- Cox Regression is more robust than parametric methods
- It can handle both discrete and continuous measures of event times.

Disadvantages

- The effect of covariates on the hazard rate but leaves the baseline hazard rate unspecified
- Does NOT assume knowledge of absolute risk [14].

2. Regularized Cox

Coxnet is a function which fits the Cox Model and is used for choosing a small number of covariates to include in the model.Because the Cox Model is rarely used for actual prediction, but we will focus on finding and interpreting an appropriate model in a rather time. In this, we have some more methods which are listed as [15]:

- Lasso-Cox
- Ridge-Cox
- EN-Cox
- OSCAR-Cox

3. CoxBoost

CoxBoost is used to fit a Cox proportional hazards model by component wise likelihood based boosting. It is especially suited for models with a large number of predictors and allows for mandatory covariates with unrealized parameter estimates [16].

4. Time-Dependent Cox

The Cox proportional-hazards regression model has achieved better use in the analysis of time-to-event data with censoring and covariates. The interrelationships between the outcome and variable over time can lead to bias if we don't understand the relationships. The form of a time-dependent covariate is much more complex than Cox models with fixed (non-time-dependent) covariates.

Advantages

- It is simple and cannot look into the future. A covariate may change based on the past data or outcomes.
- One of the strengths of the Cox model is its ability to hold within covariates that change over time [17].

5.3 Parametric Methods

Parametric methods are more efficient and accurate for estimation when the timeto the event of interest follows a particular distribution specified in terms of certainparameters. [7].

1. Linear Regression

Linear Regression is used in problems where the label is of continuous nature e.g. Sales of a retail chain. It consists of ordinary least squares method for fitting the best line that minimizes the sum of squared errors between the predicted and actual data points.

Analysis:This best fit line is known as regression line and represented by a linear equation Y= a *X + b.

In this equation:

- Y – Dependent Variable
- a – Slope
- X – Independent variable
- b – Intercept

Advantage: (Irrelevant attributes tend to get a coefficient close to zero in the hyper plane's equation.) Hypothesis function is easy to understand.

Disadvantages: Oversimplifies the classification rule. Difficult to compute. Limited to numeric attributes. Doesn't at all represent how humans learn[5].

2. Accelerated Failure Time

In survival analysis, an accelerated failure time model (AFT model) is a parametric model that provides an alternative to the commonly used proportional hazards models. Whereas a proportional hazards model assumes that the effect of a covariate is to multiply the hazard by some constant, an AFT model assumes that the effect of a covariate is to accelerate or decelerate the life course of a disease by some constant.

Advantages

- Unlike proportional hazards models, the regression parameter estimates from AFT models are robust to omitted covariates.
- They are also less affected by the choice of probability distribution. The results of AFT models are easily interpreted [23].

Advantage of Parametric Methods

- You don't need much data that could be converted in some order or format of ranks.
- Easy to interpret, more efficient and accurate when the survival times follow a particular distribution.

Disadvantage

- When the distribution assumption is violated, it may be inconsistent and can give sub-optimal results.
- The size of sample is always very big [25].

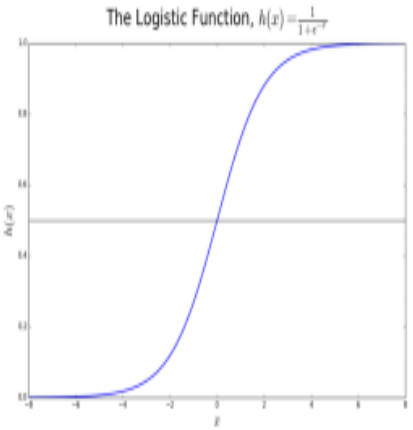
6. LIST OF COMMON MACHINE LEARNING METHODS

Here are the some of commonly used machine learning algorithms which are:

- Logistic Regression
- Survival Trees
- Bayesian Methods
- Neural Networks
- Support Vector Machines
- Advanced Machine Learning

6.1 Logistic Regression

In statistics, logistic regressionis a regression model where the dependent variable (DV) is categorical, binary dependent variable and where the output can take only two values, "0" and "1", which represent outcomes such as pass/fail, win/lose, or healthy/sick. In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as logit regression.



The logistic regression equation P(x) = e ^ (b0 +b1*x) / (1 + e^(b0 + b1*x)) can be transformed into ln(p(x) / 1-p(x)) = b0 + b1*x.

Goals: The goal of logistic regression is to use the training data to find the values of coefficients b0 and b1 such that it will minimize the error between the predicted outcome and the actual outcome. These coefficients are estimated using the technique of Maximum Likelihood Estimation [33].

Advantage:Outputs have a nice probabilistic interpretation, and the algorithm can be regularized to avoid over fitting. Logistic models can be updated easily with new data using stochastic gradient descent. It is more robust [29].

Disadvantages

- Restrictions on the Dependent Variable.
- Large Sample Size.Assumption of Linearity.
- Only for Between-Subject Designs.

6.2 Survival Trees

Survival trees are one form of classification and regression trees which are trained to handle censored data. The primary difference between a survival tree and the standard decision tree is in the choice of splitting criterion.The survival trees treat more complex situations such as those involving multivariate and correlated survival data in different directions[18].

Advantages

- The main improvement of a survival tree over the standard decision tree is its ability to handle the censored data using the tree structure.
- However, an ensemble of trees can avoid the problem of final tree selection with better performance compared to a single tree [7].

6.3 Bayesian Methods

Bayes theorem is one of the most fundamental principles in probability theory and mathematical statistics; it provides a link between the posterior probability and the priorprobability, so that one can see the changes in probability values before and after accounting for a certain event. Using the Bayes theorem, there are two models, namely [7],

- Na¨ıve Bayes (NB) and
- Bayesian network (BN)

Advantages

- Bayesian analysis can be more robust to outliers, by using more flexible distributions.
- Bayesian analysis ties better into multi-step strategies (that can be automated) such as multi-armed bandits [4].

Disadvantages

- There is no correct way to choose a prior. If you do not proceed with caution, you can generate misleading results.
- It provides a high computational cost, especially in models with a large number of parameters [8].

6.4 Neural Networks

A neural network is a machine learning algorithm based on the model of a human neuron. It sends and process signals in the form of electrical and chemical signals. These neurons are connected with a special structure known as synapses. Synapses allow neurons to pass signals. We can apply neural network not only for classification. It can also apply for regression of continuous target attributes [7].

Advantages

- ANN is nonlinear model that is easy to use, understand and perform well with linear and nonlinear data.

Disadvantages

- It may compel a neural network designer to fill millions of database rows for its connections – which can consume vast amounts of computer memory and hard disk space [19].

6.5 Support Vector Machines

Support Vector Machine (SVM), a very successful supervised learning approach, is used mostly for classification and can also be modified for regression problems. It has also been successfully adapted to in survival analysis problems.SVMs are a new promising non-linear, non-parametric classification technique, which already showed good results in the medical diagnostics, optical character recognition, electric load forecasting and other fields [7].

Advantages

- SVMs is flexibility.
- SVMs provide a good out-of-sample generalization.

Disadvantages

- SVMs is the lack of transparency of results.
- The weights of the financial ratios are not constant [20].

6.6 Advanced Machine Learning

Over the past few years, more advanced machine learning methods have been developed to deal with and predict from censored data. The AML engine is intended to work with the Symantec real-time cloud-based threat intelligence to provide best-in-class protection with low false positives. Some of the common methods approached here area as follows as [7]:

- Ensemble Learning
- Active Learning
- Transfer Learning
- Multi-Task Learning

1. Ensemble Learning

Ensemble modeling is a powerful way to improve the performance of your model.Ensemble learning is a machine learning paradigm where recognition, text categorization, face recognition, computer-aided medical diagnosis, gene expression multiple learners are trained to solve the same problem. Ensemble learning has already been used in diverse applications such as optical character analysis, etc. Some commonly used Ensemble learning techniques are [21]:

- Bagging
- Boosting
- Stacking

2. Active Learning

Active learning is a special case of semi-supervised machine learning in which a learning algorithm is able to interactively query the user to obtain the desired outputs at new data points.This type of iterative supervised learning is called active learning. Pedestrian detection, twitter negative tweet recognition are the some real world examples of active learning [7].

3. Transfer Learning

It is proposed to improve the prediction performance of the Cox model in the target domain through knowledge transfer from the source domain in the context of survival models built on multiple high-dimensional datasets. Thus, the model will not only select important features but will also learn a shared representation across source and target domains to improve the model performance on the target task [7].

4. Multi-Task Learning

This method is a generalization of the well-known single-task 1-norm regularization. We prove that the method is equivalent to solving a convex optimization problem for which there is an iterative algorithm which converges to an optimal solution [22].

7. CONCLUSION

The foremost target of ML researchers is to design more efficient (in terms of both time and space) and practical general purpose learning methods that can perform better over a widespread domain. Higher accuracy of prediction and humanly interpretable prediction rules are also of high importance. Being completely data-driven and having the ability to examine a large amount of data in smaller intervals of time, an ML algorithm has an edge over manual or direct programming. Also they are often more accurate and not prone to human bias. Likewise, it will help reform Statistical rules, by providing more computational stance. Obviously, both Statistics and Computer Science will also embellish ML as they develop and contribute more advanced theories to modify the way of learning.

8. REFERENCES

[1] Kajaree Das, Rabi Narayan Behera, "Machine Learning: A Survey, International Journal of Innovative Research (IJIR), and ISSN:2320-9801, Vol. 5, Issue 2, 2017.

[2] Sunpreet Kaur, Sonika Jindal, "Machine Learning: A Survey, International Journal of Innovative Research in Advanced Engineering (IJIRAE), ISSN: 2349-2763, Issue 11, Volume 3, 2016.
[3] Alexandra L'Heureux, Katarina Grolinger, and Miriam A. M. Capretz, Machine Learning With Big Data: Challenges and Approaches: IEEE Access, NSERC CRD, 2169-3536, Volume5, 2017.
[4] https://www.quora.com/What-are-the-advantages-of-Bayesian-methods-over-frequentist-methods-in-web-data
[5] Carl Burch, A survey of machine learning: Pennsylvania Governor's School for the Sciences, 2001.
[6] Alex Berson, Stephen Smith, and Kurt Thearling, An Overview of Data Mining Techniques: Building Data Mining Applications for CRM.
[7] Ping Wang, Yan Li, Chandan K. Reddy, Machine Learning for Survival Analysis: A Survey, ACM Computing Surveys, Vol. 1, No. 1, Article 1, Publication date: March 2017.
[8] https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introbayes_sect006.htm
[9] Nelson–Aalen Estimator Cited in http://www.medicine.mcgill.ca/epidemiology/hanley/c609/Material/NelsonAalenEstimator.pdf
[10] Hakizimana Leopord, Dr. Wilson Kipruto Cheruiyot,
[11] Dr. Stephen Kimani, A Survey and Analysis on Classification and Regression Data Mining Techniques for Diseases Outbreak Prediction in Datasets: The International Journal Of Engineering And Science (IJES), Volume 5, Issue 9, Pages PP -01-11, 2016, ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805.
[12] Nelson–Aalen estimator: Non-parametric estimation for basic survival models cited in http://www.steinsaltz.me.uk/survival/Lecture%20notes/Survival_Lecture_Notes_part2.pdf
[13] Life-Table Analysis: NCSS Statistical Software, NCSS, LLC.
[14] Cox (proportional Hazards) Regression: StatsDirect Limited, 2000-2018.
[15] Kristin Sainani, Introduction to Cox Regression: Stanford University Department of Health Research and Policy at http://www.stanford.edu/~kcobb.
[16] Noah Simon, Jerome Friedman, Trevor Hastie, Rob Tibshirani, Coxnet: Regularized Cox Regression, September 21, 2017.
[17] Harald Binder, Cox models by likelihood based boosting for a single survival endpoint or competing risks: CRAN, 2013-05-10 10:18:16.
[18] Terry Therneau, Cynthia Crowson, Elizabeth Atkinson, Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model: Mayo Clinic, April 4, 2017.
[19] Imad Bou-Hamad, Denis Larocque, Hatem Ben-Ameur, A review of survival trees: Statistics Surveys, Vol. 5 (2011) 44–71 ISSN: 1935-7516 DOI: 10.1214/09-SS047.
[20] Shailna P, Artificial Neural Network (ANN) in Machine Learning in Machine Learning Tutorials: Data Flair, 15 Jul, 2017.
[21] Laura Auria, Rouslan A. Moro, Support Vector Machines (SVM) as a Technique for Solvency Analysis: DIW Berlin German Institute for Economic Research, Berlin, August 2008.
[22] Zhi-Hua Zhou, Ensemble Learning: National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China.
[23] Andreas Argyriou, Theodoros Evgeniou, Massimiliano Pontil, Multi-Task Feature Learning: PASCAL Network of Excellence, IST-2002- 506778.
[24] https://en.wikipedia.org/wiki/Accelerated_failure_time_model
[25] https://en.wikipedia.org/wiki/Kaplan%E2%80%93Meier_estimator
[26] https://content.wisestep.com/advantages-disadvantages-parametric-tests/
[27] https://stats.stackexchange.com/questions/56500/what-are-the-main-differences-between-k-means-and-k-nearest-neighbours
[28] https://machinelearningmastery.com/parametric-and-nonparametric-machine-learning-algorithms/
[29] https://en.wikipedia.org/wiki/Machine_learning
[30] http://www.ehow.co.uk/info_8574447_disadvantages-logistic-regression.html
[31] https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/