

# A Study of Polynomial Regression towards Machine Learning

Prema Kumari<sup>1\*</sup> Dr. Aswini Kumar<sup>2</sup>

<sup>1</sup> Research Scholar, OPJS University, Churu, Rajasthan

<sup>2</sup> Associate Professor, Mathematics, OPJS University, Churu, Rajasthan

**Abstract** – In the study, we address the errand of polynomial regression, i.e., prompting regression models dependent on polynomial equations, from information. We go for enhancing and stretching out the current approaches to learning polynomial regression models in a few headings. First, we enhance the current methods for tending to the issue of over-fitting and enhance the current methods for requesting the hunt space of competitor polynomial equations. Second, we expand the extension of existing methods towards learning piecewise, multi-target, and classification through regression polynomial models. We likewise guess that their execution will be equivalent to the execution of models got with other best in class regression and classification approaches. To achieve the points and test the speculations, we begin with playing out a study of existing exploration on learning regression models with spotlight on assessment metrics utilized for regression. At that point we grow new heuristics and refinement administrators, and execute them into the algorithm Ciper for prompting polynomial regression models. The algorithm is fit for learning piecewise and multi-target polynomial models and polynomial models for classification by means of regression. At long last, we perform observational assessment and near examination of the execution of polynomial models acquired with Ciper and the execution of models got with different approaches. The consequences of the exact assessment and the relative investigation demonstrate that the recently created pursuit heuristics and refinement administrators prompt enhanced execution of the educated regression models. The execution of models induced with Ciper is equivalent to the execution of models induced with other ordinarily utilized regression algorithms. Likewise, classification models dependent on multi-target polynomials have prescient execution tantamount to the execution of models got with other classification approaches. At long last, we additionally demonstrate that piecewise polynomial models of constrained degree perform equivalent to polynomial models of higher (boundless) degrees.

**Keywords:** Polynomial Regression, Polynomial Equations, Information, Algorithm, etc.

-----X-----

## INTRODUCTION

Regression models foresee the estimation of a dependent numeric variable from the estimations of independent variables, likewise alluded to as indicators (in measurements, indicators are additionally alluded to as regressors). The regression undertaking is the issue of instigating or learning a regression model from a table of estimated estimations of the dependent and independent variables. The least difficult way to deal with the regression errand is linear regression, where the dependent variable is modeled as a linear mix of the indicators. Further developed regression approaches and models incorporate regression and model trees and in addition multivariate adaptive regression splines (MARS).

This study manages the assignment of polynomial regression, i.e., the errand of initiating a regression model as a polynomial equation that predicts the

estimation of a dependent numeric variable. We expand upon a current way to deal with polynomial regression, Ciper . Ciper performs heuristic pursuit through the space of hopeful polynomial equations beginning with the least difficult polynomial and adding terms to it at each progression of the inquiry to touch base at more unpredictable ones. Every competitor structure is coordinated against preparing information and estimations of consistent parameters are gotten that prompt the maximal fit to the information.

In any case, utilizing just level of fit to control the hunt is certifiably not a smart thought, since it unquestionably prompts over-fitting the preparation information . Note that polynomial models can consummately add any information, since it is realized that any  $n$  focuses can be splendidly added with a  $(n-1)$ -th degree polynomial. To address this issue, Ciper joins the level of fit with the polynomial's intricacy to control the heuristic pursuit.

We address a few restrictions of the first form of Ciper in this study. First, Ciper utilizes a specially appointed weighting plan to join level of fit with the polynomial's multifaceted nature to acquire the estimation of the heuristic capacity for a given applicant polynomial. Conversely, we perform here a top to bottom investigation of various methodologies to battle the issue of over-fitting with polynomial regression models. Second, Ciper utilizes a straightforward easy to-complex requesting of the hunt space that, joined with a particular heuristic capacity, may prompt under-looking through the space of hopeful models. In the study, we investigate distinctive refinement administrators for requesting the space of polynomial models. Third, Ciper centers around learning a polynomial model that predicts the estimation of a solitary dependent variable and is legitimate over the entire preparing dataset. Here, we expand the extent of polynomial regression toward multi-target regression models that can all the while foresee a few dependent variables.

Additionally, we create approaches to learning piecewise polynomial models. At long last, we utilize multitarget polynomial models on classification undertaking by applying the classification by means of regression approach.

The Polynomial Regression strategy is intended to develop a statistical model portraying the effect of a solitary quantitative factor  $X$  on a dependent variable  $Y$ . A polynomial model including  $X$  and forces of  $X$  is fit to the information. Tests are hurried to decide the best possible request of the polynomial. The fitted model might be plotted with certainty limits as well as expectation limits. Residuals may likewise be plotted and persuasive perceptions distinguished.

Polynomial regression is a sort of linear regression in which the association between the input variables  $x$  and the output variable  $y$  is modeled as a polynomial. But polynomial regression fits a nonlinear model to the data, as a statistical estimation issue it is linear, as in the regression work is linear in the dark parameters that are evaluated from the data. Hence, polynomial regression is seen as a phenomenal case of linear regression.

## MACHINE LEARNING:

Machine learning is a part of man-made reasoning, worried about the structure and advancement of algorithms that can enhance their conduct dependent on exact information. The experimental information take a type of precedents that delineate relations between watched variables. A noteworthy focal point of machine learning research is to naturally figure out how to perceive designs in the models and settle on clever choices.

A huge piece of machine learning manages the errand of modeling, i.e., building prescient models.

This models foresee the estimation of a dependent variable from the estimations of independent variables, additionally alluded to as indicators. Prescient modeling issues can be partitioned into classification and regression issues.

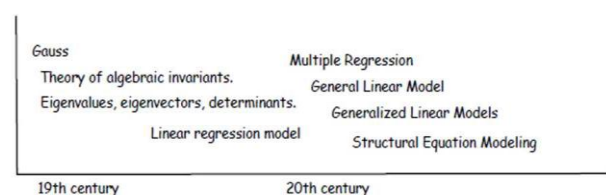
Classification issues include anticipating the estimations of an all out (ostensible) output variable. At least one persistent or all out input variables can be utilized as indicators. There are various methods for taking care of classification issues that include straightforward nonstop indicators, all out indicators, or both.

Regression issues include foreseeing the estimation of a constant variable from at least one nonstop or all out variables. For instance, one might need to anticipate the offering cost of a solitary family home from different ceaseless variables and absolute (ostensible) variables. Multiple regression can be connected for this issue, to locate a linear equation that can be utilized to foresee the offering costs from alternate variables.

Inside machine learning, various progressed statistical methods exist for taking care of regression and classification undertakings with multiple input variables and (normally) a solitary output variable. These methods incorporate Support Vector Machines (SVM) for classification and regression, Naive Bayes for classification, k-Nearest Neighbors (KNN) for classification and regression, Classification and Regression Trees (CART), Multivariate Adaptive Regression Splines (MARSplines), and others. Huge group of regression methods is the class of general linear regression methods, depicted beneath.

## GENERAL LINEAR REGRESSION:

The foundations of regression investigation return to the beginnings of arithmetic. The theory of arithmetical invariants created from crafted by nineteenth century mathematicians, for example, Gauss, Boole, Cayley and Sylvester made the linear regression model conceivable. The theory recognizes those amounts in frameworks of equations that stay unaltered under linear changes of the variables in the framework. A portion of the new ideas presented by this theory are eigenvalues, eigenvectors, determinants, and framework deterioration methods.



The theory was before long stretched out to the linear regression model and relationship methods.

They fill in as the reason for the improvement of the general linear model. The general linear model can be viewed as an augmentation of linear multiple regression for a solitary output variable .

**Multiple Regression** - The general pm posture of multiple regression<sup>1</sup> is to measure the connection between a few input variables and an output variable. It is expected that the output (dependent) variable y is linearly identified with the input (independent, indicator) variables  $X_1, \dots, X_p$  as below,

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon = \beta_0 + \sum_{i=1}^p X_i \cdot \beta_i + \varepsilon \quad (1)$$

Where  $\varepsilon$  is an inconspicuous arbitrary variable (the error segment) with mean 0 and difference  $\sigma^2$ . The relationship portrayed by Equation 1 is known as a linear regression model, where  $\beta_0, \beta_1, \dots, \beta_p$  are obscure parameters and  $\sigma^2 > 0$  is an obscure error fluctuation. The linearity of the model is an aftereffect of its linearity in the parameters  $\beta_0, \beta_1, \dots, \beta_p$ . Transformations of the input variables, (for example, powers  $X_j^d$  and items  $X_j \cdot X_k$ ) can be incorporated into the model without it losing its portrayal as a linear regression model. The regression coefficients  $\beta_0, \beta_1, \dots, \beta_p$  represent the independent contributions of each input variable to the forecast of the output variable.

Normally, the parameters  $\beta$  are evaluated from an arrangement of preparing information  $(x_1, y_1), \dots, (x_N, y_N)$ . Each  $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})$  is a vector of highlight estimations for the i-th case. The most prominent estimation method is slightest squares, in which the coefficients  $\beta_0, \beta_1, \dots, \beta_p$  limit the leftover total of squares

$$RSS(\beta) = \sum_{i=1}^N (y_i - (\beta_0 + \sum_{j=1}^p x_{i,j} \cdot \beta_j))^2 \quad (2)$$

Mean by X the  $N \times (p+1)$  matrix with each row an input vector (with a 1 in the first position,  $X = (1, X_1, \dots, X_p)$ ). Similarly, let  $y = (y_0, y_1, \dots, y_N)$  be the N dimensional vector of outputs in the preparation set. The equation 1 can be rewritten as pursues:

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,p} \\ 1 & x_{2,1} & \dots & x_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N,1} & \dots & x_{N,p} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (3)$$

where  $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_N)$  is the vector of errors/residuals  $\varepsilon$ .

The residual sum of squares is then:

$$RSS(\beta) = (y - X\beta)^T \cdot (y - X\beta). \quad (4)$$

Assuming that X has full column rank, and hence  $X^T X$  is positive definite, by setting the first derivative to zero

$$X^T \cdot (y - X\beta) = 0 \quad (5)$$

the unique solution to the minimization problem defined by Equation 2 is found to be:

$$\hat{\beta} = (X^T X)^{-1} \cdot X^T y. \quad (6)$$

The variance of residuals  $\sigma^2$  is estimated using the equation:

$$\hat{\sigma}^2 = \frac{1}{(N - p - 1)} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (7)$$

where  $\hat{y}_i$  is the predicted value of y at  $x_i$ .

The multiple regression model can be utilized to dissect just a solitary output variable. It cannot give an answer for the regression coefficients when the independent variables X are linearly dependent and the converse of  $X^T X$  does not exist. Diverse approaches introduced beneath can be utilized to address these issues.

**Multiple Output Variables** - The general linear model can deal with a few output variables without a moment's delay. The y vector of N perceptions of a solitary variable can be supplanted by a Y lattice of N perceptions of m distinctive Y variables. Correspondingly, the  $\beta$  vector of regression coefficients for a solitary Y variable can be supplanted by a  $\beta$  network of regression coefficients, with one vector of  $\beta$  coefficients for every one of the m output variables. These substitutions yield what is in some cases called the multivariate regression model, yet it ought to be underlined that the framework details of the multiple and multivariate regression models are indistinguishable, with the exception of the quantity of segments in the Y and  $\beta$  networks. The method for unraveling for the  $\beta$  coefficients is additionally indistinguishable, that is, m distinctive arrangements of regression coefficients are independently found for the m diverse output variables in the multivariate regression model.

The general linear model can give an answer for the Equation 2 when the input variables are linearly dependent and in this way the reverse of  $X^T X$  does

not exist. A non-full-rank framework doesn't have a standard opposite. This issue is settled in the general linear model by utilizing a generalized backwards of the  $X^T X$  framework. One method for doing this is to utilize regularization approaches like in edge regression that punishes the size of the  $\beta$  coefficients. The edge regression arrangements are given by the accompanying equations:

$$\beta_\lambda = (X^T X + \lambda I)^{-1} \cdot X^T y \quad (8)$$

where  $\lambda$  controls the measure of punishment identified with the greatness of the coefficients.

**All out Variables** - The general linear model is often connected to investigate information that has all out (ostensible) input variables. For instance, sexual orientation is plainly a straight out dimension variable. There are two essential methods by which sexual orientation can be coded into at least one input variables: the sigma-confined method and the over parameterized method.

Utilizing the sigma-limited method, the guys are relegated with the esteem - 1 and the females are doled out with the esteem 1. The qualities on the subsequent input variable, 1 and - 1, speak to a quantitative complexity among guys and females. On the off chance that the regression coefficient for the variable is certain, the gathering coded as 1 on the input variable will have a higher anticipated an incentive on the output variable, and if the regression coefficient is negative, the gathering coded as - 1 on the input variable will have a higher anticipated an incentive on the output variable. The sigma-confined parametrization of clear cut input variables for the most part prompts  $X^T X$  matrices which don't require a generalized opposite for tackling the minimization issue characterized by Equation 2.

The over parameterized method for recoding absolute indicators is the pointer variable methodology. In this method, a different input variable is coded for each gathering distinguished by a downright input variable. For instance, females may be doled out an estimation of 1 and guys an estimation of 0 on a first input variable distinguishing enrollment in the female sexual orientation gathering. Guys would then be doled out an estimation of 1 and females an estimation of 0 on a second input variable distinguishing participation in the male sex gathering. This method of recoding clear cut variables will dependably prompt  $X^T X$  networks with repetitive sections. Consequently, it requires a generalized reverse for tackling the minimization issue characterized by Equation 2.

**Generalized Linear Models** - There are numerous connections that can't be portrayed by a linear

equation. There are two noteworthy explanations behind this.

The first reason is the circulation of the output variable. The output variable of intrigue may have a noncontiguous circulation, and in this way, the anticipated qualities ought to likewise pursue the separate dissemination. For instance, we might be occupied with foreseeing one of three conceivable discrete results. The output variable can just interpretation of 3 unmistakable qualities, and the dissemination of the output variable is said to be multinomial. Or on the other hand guess we are endeavoring to anticipate what number of kids families will have, as a component of salary and different other financial pointers. The output variable number of youngsters is discrete, and no doubt the dispersion of that variable is exceptionally skewed (i.e., most families have 1, 2, or 3 kids, less will have 4 or 5, not very many will have 6 or 7, et cetera). For this situation, it is sensible to expect that the output variable pursues a Poisson dispersion.

The second reason, why the linear model may be lacking to portray a specific relationship, is that the impact of the indicators on the output variable may not be linear in nature. For instance, the connection between a man's age and different markers of wellbeing is in all likelihood not linear. The generalized linear model can be utilized to foresee reactions both for output variables with discrete appropriations and for output variables which are nonlinearly identified with the indicators with a connection work,

In the generalized linear model, the connection between the output variable  $y$  and the input variables  $X$  is thought to be

$$y = g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_p) + \varepsilon \quad (9)$$

where  $g$  is a function. The inverse function of  $g$ , say  $f$ , is called the *link function*.

$$f(\hat{y}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_p + \varepsilon \quad (10)$$

where  $\hat{y}$  stands for the expected value of  $y$ . Various link functions can be chosen, depending on the assumed distribution of the  $y$  variable:

- Identity link:  $f(z) = z$
- Log link:  $f(z) = \log(z)$
- Power link:  $f(z) = z^a$ , for a given  $a$



- Logit link:  $f(z) = \log(z/(1-z))$ .

The parameters  $\beta$  are usually estimated by maximum likelihood estimation, which requires the use of iterative computational procedures.

### **Building Generalized Linear Models on Subsets of Predictors-**

When building generalized linear models notwithstanding fitting a model of the predetermined kind utilizing every single accessible indicator, diverse methods for programmed model building can be employed that select the utilized indicators in various ways. For the particular sort of model close by, to assemble models on subsets of indicators, we can utilize diverse methods for automatic model building. They include: forward selection, in reverse evacuation, forward stepwise, in reverse stepwise techniques, and best-subset look systems. In forward methods of selection of impacts (variables) to incorporate into the model, score insights are contrasted with select new noteworthy impacts.

Stepwise regression systems include recognizing an underlying model, more than once changing the model at the past advance by including or evacuating an input variable as per the venturing criteria, and ending the hunt while venturing is never again conceivable given the venturing criteria. For the forward stepwise and forward passage methods, the underlying model dependably incorporates the regression catch. The underlying model may incorporate at least one impacts indicated to be constrained into the model.

In best-subset regression, the quantity of conceivable sub models increments quickly as the quantity of impacts (variables) incorporated into the model increments. The measure of calculation required to play out every single conceivable subset regression increments as the quantity of conceivable submodels increments, and holding all else steady, likewise increments quickly as the quantity of levels for impacts including all out indicators expands, subsequently bringing about more sections in the structure grid X. Every conceivable subset of up to twelve or so impacts could surely hypothetically be registered for a structure that incorporates two dozen or so impacts, all of which have numerous dimensions, however the calculation would be exceptionally tedious.

### **MODELING USING POLYNOMIAL REGRESSION:**

Regression examination includes distinguishing the connection between a dependent variable and at least one independent variables. It is a standout amongst the most imperative statistical instruments which is widely utilized in all sciences. It is

exceptionally utilized in business and financial matters to think about the connection between at least two variables that are connected causally. A model of the relationship is hypothesized, and gauges of the parameter esteems are utilized to build up an expected regression equation.

Different tests are then utilized to decide whether the model is acceptable. Model approval is an essential advance in the modeling procedure and aides in evaluating the dependability of models before they can be utilized in basic leadership.

### **The multiple regressions -**

Multiple regressions alludes to regression applications in which there are in excess of one independent variables. Multiple regression incorporates a method called polynomial regression. In polynomial regression we relapse a dependent variable on forces of the independent variables.

#### **1. The multiple regression model**

The essential multiple regression model of a dependent (reaction) variable Y on an arrangement of k independent (indicator) variables  $X_1, X_2, \dots, X_k$  can be communicated as

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \dots + \beta_k x_{1k} + e_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \dots + \beta_k x_{2k} + e_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \dots + \beta_k x_{nk} + e_n \end{cases} \quad (11)$$

i.e.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i, \text{ for } i = 1, 2, \dots, n \quad (12)$$

Where  $y_i$  is the estimation of the dependent variable Y for the ith case,  $x_{ij}$  is the estimation of the y'th independent variable  $X_j$  for the ith case,  $\beta_0$  is the T-block of the regression surface (think multidimensionality), each  $\beta_j, j = 1, 2, \dots, k$ , is the slant of the regression surface as for variable  $X_j$  and  $e_i$  is the irregular error segment for the ith case. In fundamental equations (11) we have n perceptions and k indicators ( $n > k + 1$ ).

The suspicions of the multiple regression model are like those for the straightforward linear regression model. Model presumptions :

- For every perception the errors  $e_i$  are regularly conveyed with mean zero and standard deviation  $\sigma$  and are independent of

the error terms related with every other perception. The errors are uncorrelated with one another. That is  $e_i \sim N(0, \sigma^2)$  for all  $i = 1, 2, \dots, n$ , independent of different errors.

- In the context of regression examination, the variables  $X_j$  are viewed as settled amounts, in spite of the fact that in the setting of relationship investigation, they are arbitrary variables. Regardless,  $X_j$  are independent of the error term. When we expect that  $X_j$  are settled amounts, we are accepting that we have acknowledge of  $k$  variables  $X_j$  and that the main irregularity in  $Y$  originates from the error term.

In grid documentation, we can rework model (1) as

$$Y = X\beta + e \quad (13)$$

where reaction vector  $Y$  and error vector  $e$  are segment vectors of length  $n$ , vector of parameters  $\beta$  is section vector of length  $k + 1$  and structure network  $X$  is  $n$  by  $k + 1$  framework (with its first segment having all components equivalent to 1, the second segment being filled by the observed estimations of  $X$ ), and so on.). We need to appraise obscure estimations of  $\beta$  and  $e$ .

## 2. Slightest squared error approach in grid shape

We gauge the regression parameters by the method of slightest squares. This is an expansion of the system utilized in straightforward linear regression. First, we compute the whole of the squared errors and, second, locate an arrangement of estimators that limit the aggregate.

Utilizing equation (13) we get for the errors

$$e = Y - X\beta. \quad (14)$$

Discover estimator  $\hat{\beta}$  we need to limit the total of squares of the errors

$$e^T e = (Y - X\beta)^T (Y - X\beta) \quad (15)$$

where the image  $0^T$  signifies the transpose of the lattice.

Here  $e^T e$  is scalar. We can take the first derivate of this protest work as for the vector Making these equivalent to 0 (a vector of zeros) we acquire typical equations .

$$X^T X \hat{\beta} = X^T Y. \quad (16)$$

Increase the reverse lattice of  $(X^T X)^{-1}$  on the both left sides in equation (16), and we have the slightest squared estimator for the multiple regression model in framework shape

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (17)$$

Vector  $\hat{\beta}$  is an unprejudiced estimator of  $\beta$ . The fitted (anticipated) values for the mean of  $Y$  (given us a chance to call them  $\hat{Y}$ ), are registered by

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY \quad (18)$$

Where  $H = X(X^T X)^{-1} X^T$ . We call this the cap framework on the grounds that is transforms  $Y$  into  $\hat{Y}$ . Grid  $H$  is symmetric, i.e.  $H = H^T$  and idempotent, i.e.  $H^2 = H$ .

The fitted qualities for error terms  $e$ , are residuals  $\hat{e}_i, i = 1, 2, \dots, n$ , that are registered by

$$\hat{e} = Y - \hat{Y} = Y - HY = (I - H)Y \quad (19)$$

where  $I$  is a personality lattice.

The total of squares of the residuals  $SSE = \hat{e}^T \hat{e}$  has the  $\chi^2$  dissemination with  $df_e = n - (k + 1)$  degrees of opportunity, and is independent of  $\hat{\beta}$ .

## CONCLUSION:

In this study, we have tended to the assignment of polynomial regression, i.e., learning polynomial regression models from information. Polynomial models have been utilized broadly previously, yet they have been to a great extent overlooked by the machine learning network. As of late, a machine learning algorithm Ciper for learning polynomial equations for regression has been produced and assessed. The algorithm has turned out to be a decent student, being practically identical to model trees and beating linear and stepwise regression. Be that as it may, Ciper has a few confinements: a restricted refinement administrator, a specially appointed heuristic capacity, no support for multiple objectives, and no support for piecewise models. The primary inspiration for playing out the work inside this study was to conquer these confinements.

To this end, we have grown new methods that enhance and expand the Ciper algorithm for

polynomial regression. The upgrades incorporate new hunt space techniques (a refinement administrator) and heuristic capacities (MDL and CV based heuristics) for assessing the execution of polynomial regression models. The augmentations expand the extent of polynomial regression toward piecewise and multi-target polynomial models and enable the utilization of polynomial models to perform classification by means of regression.

Regression examination is a statistical device for the examination of connections between variables. The multiple regression investigation is a valuable method for creating numerical models where there are a few (more than two) variables included.

Polynomial regression model comprises of progressive power terms. Each model will incorporate the most elevated request term in addition to all lower arrange terms (noteworthy or not). We can see polynomial regression as a specific instance of multiple linear regression. Polynomial models are a compelling and adaptable bend fitting strategy.

## REFERENCES

1. Aha, D. (1992). Generalizing from case studies: A case study. In: Proceedings of the 9<sup>th</sup> International Workshop on Machine Learning, Morgan Kaufmann, pp. 1–10
2. Alex S. & Vishwanathan, S.V.N. (2008). *Introduction to Machine Learning*. Published by the press syndicate of the University of Cambridge, Cambridge, United Kingdom. Copyright © Cambridge University Press 2008. ISBN: 0-521 82583-0. Available at KTH website: <https://www.kth.se/social/upload/53a14887f276540ebc81aec3/online.pdf> Retrieved from website: <http://alex.smola.org/drafts/thebook.pdf>
3. Blake, C.L. & Merz, C.J. (1998). UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science.
4. Brazdil, P., Soares, C., da Costa, J. (2003). Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results. *Machine Learning* 50(3), pp. 251–277
5. Breiman L. (2001). Random Forests, *Machine Learning*, 45: pp. 5-32.
6. Breiman, L.; Spector, P. (1992). Submodel selection and evaluation in regression. *International Statistical Review* 60, pp. 291–319.
7. Chatterjee, C.; Sarkar, R. (2009). Multi-step polynomial regression method to model and forecast malaria incidence. *PLoS One* 2009, 3, pp. e4726.
8. D. Michie, D.J.S., Taylor, C.C., eds. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, New York.
9. Domingos, P. & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* Volume 29, pp. 103–130 Copyright © 1997 Kluwer Academic Publishers. Manufactured in The Netherlands. Available at University of Trento website: <http://disi.unitn.it/~p2p/RelatedWork/Matching/domingos97optimality.pdf>
10. Elder, J. (n.d). (2011). Introduction to Machine Learning and Pattern Recognition. Available at LASSONDE University EECS Department York website: [http://www.eecs.yorku.ca/course\\_archive/2011-12/F/44045327/lectures/01%20Introduction.pdf](http://www.eecs.yorku.ca/course_archive/2011-12/F/44045327/lectures/01%20Introduction.pdf)

---

## Corresponding Author

### Prema Kumari\*

Research Scholar, OPJS University, Churu, Rajasthan