JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Linear Regression Algorithm for Data Science

Ram Dulari
Student
Computer Science & Engineering at DPG Institute of Technology and Management.
Mrs. Richa Nehra
Assistant professor,
CSE Department DPG Institute of Technology and Management.

ABSTRACT

AI research has focused on improving the decision-making capabilities of computers, i.e., the ability to select high-quality actions in pursuit of a given objective. When the objective is aligned with the values of the human race, this can lead to tremendous benefits. When the objective is misaligned, improving the AI system's decision-making may lead to worse outcomes for the human race. The objectives of the proposed research are (1) to create a mathematical framework in which fundamental questions of value alignment can be investigated; (2) to develop and experiment with methods for aligning the values of a machine (whether explicitly or implicitly represented) with those of humans; (3) to understand the relationships among the degree of value alignment, the decision-making capability of the machine, and the potential loss to the human; and (4) to understand in particular the implications of the computational limitations of humans and machines for value alignment. The core of our technical approach will be a cooperative, game-theoretic extension of inverse reinforcement learning, allowing for the different action spaces of humans and machines and the varying motivations of humans; the concepts of rational meta reasoning and bounded optimality will inform our investigation of the effects of computational limitations.

ABOUT THE AUTHORS

Ram Dulari is a student of Computer Science & Engineering at DPG Institute of Technology and Management. She is currently doing her master's from this Institute and Degree is about to complete in Jun 2021. Her research area is in Artificial Intelligence, Data Science. Before her master's she is having a bachelor's degree in computer science and Engineering. And moreover, a technical software industry experience. Two years experience with NSDC Council in agriculture and food sector as a Coordinator (Government).

Mrs. Richa Nehra is an assistant professor of CSE Department DPG Institute of Technology and Management. Her research includes in software Engineering and Data Analysis, etc. She has published paper widely; She has 3 year of teaching and research experience. Her interest is in data fields. Before her master's degree she is having a bachelor's degree (B.Tech) in Computer Science and Engineering Department. She has an excellent academic record.

CHAPTER : Linear Regression

Linear Regression is a machine learning algorithm based on Line of Regular Reading Reversion enables the function to predict the variable value of a dependency (y) based on a given independent variation (x). Therefore, this regression process finds an equal relationship between x (input) and y (output).

In simple terms, Linear Regression is a controlled Installation model where the model finds a very consistent line between independent and reliable variants i.e. it finds a balanced relationship between dependent and independent variations.

Linear Regression is of two types: Simple and Multiple. Linear easy control is where there is one independent variation and the model has to find its linear relationship and dependent variations.

While, in Multiple Linear Regression there are many independent variations of the model to find a relationship.

Simple Linear Regression equation, where bo is odd, b1 equals or slopes, x independent variables and y dependent variations.

$$y = b_o + b_1 x$$

Equation of Multiple Linear Regression, where bo is the intercept, $b_1, b_2, b_3, b_4 \ldots, b_n$ are coefficients or slopes of the independent variables $x_1, x_2, x_3, x_4 \ldots, x_n$ and y is the dependent variable.

$$y = b_o + b_1 x_1 + b_2 x_2 + b_3 x_3 \ldots . + b_n x_n$$

A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized. Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.

Let's understand this with the help of a diagram.

Image Source: Statistical tools for high-throughput data analysis In the above diagram

,

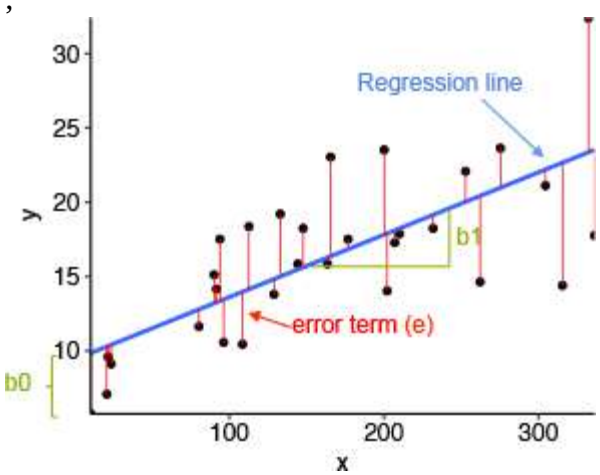


Image Source: Statistical tools for high-throughput data analysis In the above diagram

1. x is our dependent variable which is plotted on the x-axis and y is the dependent variable which is plotted on the y-axis.
2. Black dots are data points i.e. real values.
3. bo is 10 intercept and b1 is a slope of x variability.
4. The blue line is the most consistent line predicted by the model i.e. the predicted values lie in the blue line.

The vertical distance between the data point and the return line is known as error or remainder. Each data point has one residue and the sum of all the differences is known as Sum of Residuals / Errors.

Mathematical Method:

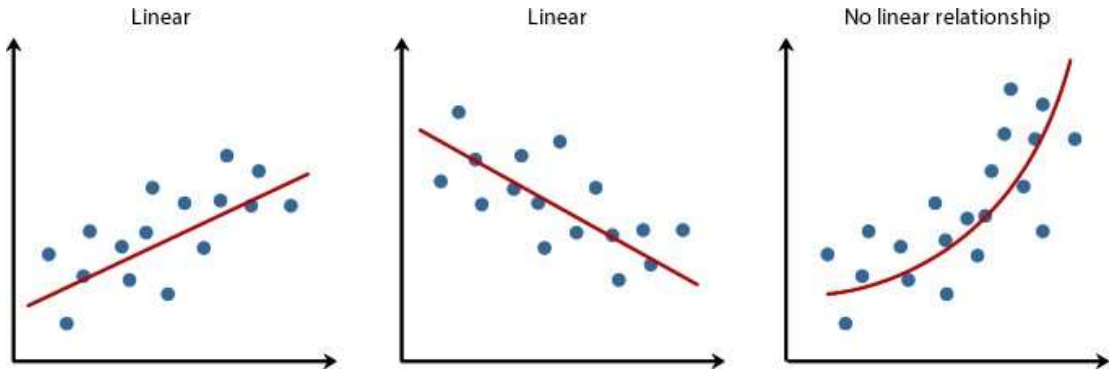
Remaining / Error = Actual Values - Estimated Values Residence / Errors Total = Total (Estimated Values) Residential Number Square / Errors = (Total (Real Estimated Values) 2

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

Assumptions of Linear Regression:
The basic assumptions of Linear Regression are as follows:

1. Linearity: It says that Y-dependent variables should be directly related to independent variables. This consideration can be tested by planning a distribution strategy between the two variables.

Benefits of Linear Regression:
Easy implementation:



Linear Regression is a very simple algorithm that can be easily used to get satisfactory results. Moreover,

these models can be easily and effectively trained even in low-power computational systems compared to other sophisticated algorithms. Line deviation has a much lower duration compared to other machine learning algorithms. Linear regression statistics are also easy to understand and interpret. So the retreat of the line is much easier to understand.

Working on classified databases:

The linear regression corresponds to the data split almost completely and is often used to determine the nature of the relationship between the variables.

In addition to the measurement can be reduced by doing so:

The linear regression corresponds to the data split almost completely and is often used to determine the nature of the relationship between the variables.
What is more appropriate is the situation in which the machine learning model fits well with the database and that is why it captures noisy data.This negatively affects the performance of the model and reduces its accuracy in the test set.
Redesign is an easy-to-use method that can effectively reduce the workload in order to reduce the risk of overcrowding.

Disadvantages of queuing:

Less than ideal: A situation that arises when a machine learning model fails to enter data correctly. This is especially the case when the hypothesis function cannot properly measure the data.

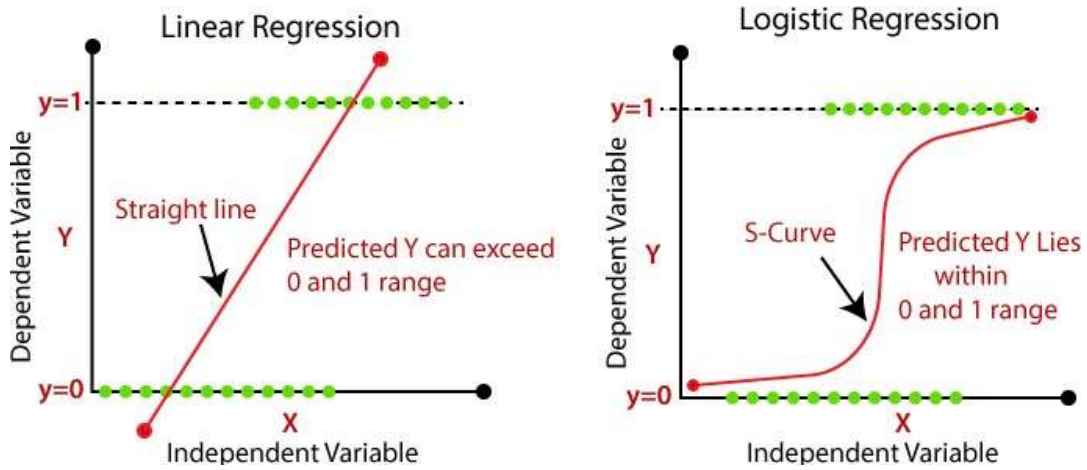
Sensitivity of exporters: Outputs to the data set are inconsistent or exaggerated values deviate from other distribution data points.The data transactions may severely compromise the performance of the machine learning model and may result in models with lower accuracy.

Outliers can have a very significant impact on the performance of line order and should therefore be dealt with properly before line deployment is used in the database.

Weight loss is used to determine the degree of difficulty before more than one definition. The process is exactly the same as several queues, except that the flexibility of the response is smooth. The result is the effect of each variable variable in the scale of the perceived event of interest.

Direct Transformation vs Postponement of Items

Linear Regression and Logistic Regression are the two most popular machine learning methods under the supervised learning method. As both algorithms are monitored in the environment this is why these algorithms use a labeled dataset to make predictions. But the main difference between them is how they are used. Linear Regression is used to solve Regression problems and Logistic Regression is used to solve Separation problems. A description of both algorithms is provided below and the difference table.



Conclusion

Machine learning can be monitored or unsupervised. If you have small data and a clear label training label, select Supervised Reading. Unattended reading can provide better performance and results for larger data sets. If you have a large set of easily accessible data, go with in- depth reading strategies. He also studied Certified Certified Reading and Reading. Now you know what Neural Networks are, their uses and limitations.

Finally, when it comes to building your own learning machine models, look at the various language development options, IDEs and Platforms. The next thing you need to do is start learning and mastering each machine learning process. The title is great, it means there is a wide range, but if you look at the depth, each article can be read in a few hours. Each topic is unique. You need to look at one topic at a time, read it, create it and apply algorithm / s to it using the language of your choice. This is a great way to start studying Learning Machines. Familiarizing yourself with one topic at a time, you will soon find the required range at the end of a typewriter.

References

[1] Leo Breiman, Random Forests. Machine Learning, Volume 45, Issue 1, October 1 2001, Pages 5-32

[2] Jerome H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, Vol. 29, No. 5 (Oct., 2001), pp. 1189-1232

[3] Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning: Data mining, inference, and prediction. Second Edition (2009), Springer Series in Statistics, Pages 219-230

[4] Georg Heinze and Michael Schemper, A solution to the problem of separation in logistic regression, Statistics in Medicine, Volume 21, Issue 16 (30 August 2002), Pages 2409–2419.

[5] Andrew Gelman et al, A weakly informative default prior distribution for logistic and other regression models, The Annals of Applied Statistics 2008, Vol. 2, No. 4, Pages 1360– 1383

[6] Lawrence R. Rabiner, A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE, Vol. 77, No. 2. (06 February 1989), Pages. 257- 286

[7] Charles C Holt, Forecasting Trends and Seasonal by Exponentially Weighted Averages. International Journal of Forecasting, Volume 20, Issue 1 (January–March 2004), Pages 5–10.

[8] M. Xie et al, A seasonal ARIMA model with exogenous variables for Elspot electricity prices in Sweden, 2013 10th International Conference on the European Energy Market (EEM), Stockholm (May 2013), Pages 1-4

[9] Cristina Stolojescu-Crisan, Data mining based wireless network traffic forecasting, 2012 10th International Symposium on Electronics and Telecommunications (ISETC), Timisoara (Nov 2012), Pages

115-118.
[10] Josef Kittler et al, On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3 (March 1998), Pages 226-239.
[11] Papadopouli M, Evaluation of short-term traffic forecasting algorithms in wireless networks, 2006 2nd Conference on Next Generation Internet Design and Engineering, NGI,
Valencia (April 2006), Pages 102-109
[12] Yantai Shu et al, Wireless traffic modeling and prediction using seasonal ARIMA models, IEEE International Conference on Communications, 2003 (ICC 2003), Anchorage
[13] Artificial intelligence: Alex is the founder of InnoArchiTech and InnoArchiTech Institute, as well as the author of *AI for People and Business* published by O'Reilly Media..
[14] Machine Learning: This article was originally published on The Conversation
[15] F.Caron and Y. W. Teh. "Bayesian Nonparametric Models for Ranked Data". In:Proceedings of the 26th Annual Conference on Neural Information Process-ing Systems (NIPS). 2012, pp. 1529–1537.
[16] U. Chajewska, D. Koller, and D. Ormoneit. "Learning an Agent's Utility Func-tion by Observing Behavior". In:Proceedings of the 18th International Con-ference on Machine Learning (ICML). 2001, pp. 35–42.
[17] M. Babe-Vroman, V. Marivate, K. Subramanian, and M. L. Littman. "Ap-prenticeship Learning About Multiple Intentions". In:Proceedings of the 28thInternational Conference on Machine Learning (ICML). 2011, pp. 897–904.