

UNIVERSIDAD PANAMERICANA

Renata Quintero del Toro
Jorge Luis Cárdenas Gómez
Marco Antonio Mendoza López

Simulation and Data Visualization
Professor Gabriel Castillo Cortés
Final Project: ML Algorithms

May 28th, 2025

Machine Learning Algorithms: Logistic Regression, Multilinear Regression & Polynomial Regression

1. Introduction

Machine learning utilizes regression algorithms to design the relationship between input variables and continuous responses. Logistic Regression, Multilinear Regression, and Polynomial Regression are of particular significance because of their inherent significance and frequent application across various domains such as healthcare, finance, marketing, and engineering. The objective of this research is to examine the three algorithms through a critical review of their inherent characteristics, industrial practical applications, and a comparison analysis comprising their advantages, limitations, and suggested applications.

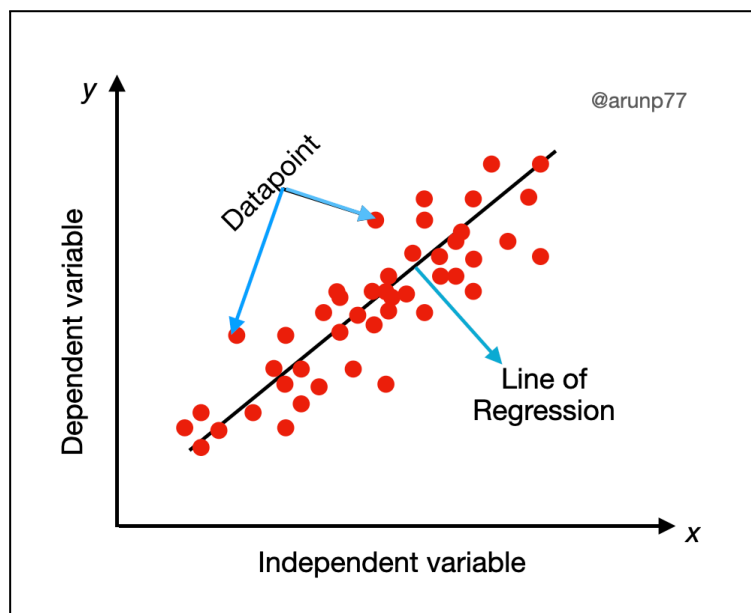


Figure 1.1 Graphical representation of a regression algorithm

2. Logistic Regression

Logistic Regression is a supervised machine learning model that is widely employed for binary classification problems. It estimates the likelihood of a particular input being in a specific class by utilizing the logistic (sigmoid) function, which plots predictions on a 0 to 1 scale. Therefore, it is ideally suited for classification issues whenever the output is categorical.

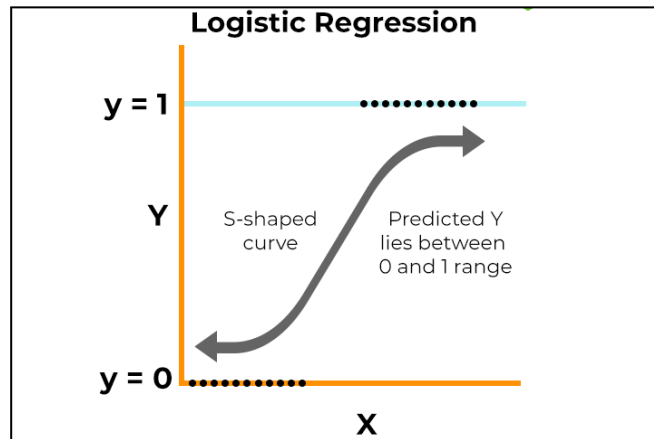


Figure 2.1 Key Assumptions for Implementing Logistic Regression

The model is very interpretable and intuitive, and the coefficients represent the influence of each input variable on the log-odds of the target. Both training and prediction are generally quick, so it is well-suited for real-time applications. Logistic Regression performs well on moderately sized datasets with clean, linearly separable data. Feature engineering, such as scaling and one-hot encoding for categorical variables, is typically necessary. The algorithm is moderately sensitive to missing and noisy data, for which preprocessing and imputation methods need to be implemented. It presumes a linear connection between the independent variables and the log-odds of the dependent variable, errors are independent, and no multicollinearity. Logistic Regression can be extended to big data using optimization methods such as stochastic gradient descent. The performance of the model is generally measured based on various metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Regularization techniques like L1 and L2 are favored to prevent overfitting and promote generalization. Logistic Regression has wide-ranging applications in credit scoring, customer churn prediction in marketing, and disease probability prediction in medical diagnosis in industry. Additionally, it is utilized as a baseline classifier in the majority of real-world tasks and research studies due to its simplicity, interpretability, and balance efficiency.

3. Multilinear Regression

Multilinear Regression, or Multiple Linear Regression, is a supervised learning approach to addressing regression issues characterized by continuous output. The technique specifies the connection between two or more independent variables and a single dependent variable through the empirical use of a linear equation. There is some degree of simplicity and interpretability of the model, with its coefficients providing a quantitative value of the impact brought about by every independent variable on the dependent variable. Both training and inference are quick, especially for small to medium datasets, and its statistical foundations make it a robust and stable approach to predictive modeling. The algorithm takes both numerical input variables, and categorical variables need encoding. Normalization or standardization is often necessary, especially if predictors have varying scales. Multilinear Regression is susceptible to missing and noisy data, so preprocessing is essential. It presumes linearity, independence, homoscedasticity, and normality of errors. The model is extendable with regularization and matrix factorization methods. It measures performance based on R-squared, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Regularization techniques such as Ridge and Lasso are available to avoid overfitting and enhance predictability. In business, Multilinear Regression is used in real estate to predict housing prices, economics to forecast demand, and engineering to predict equipment failure. Its ease of use and interpretability remain a staple of business analytics and scientific research as well.

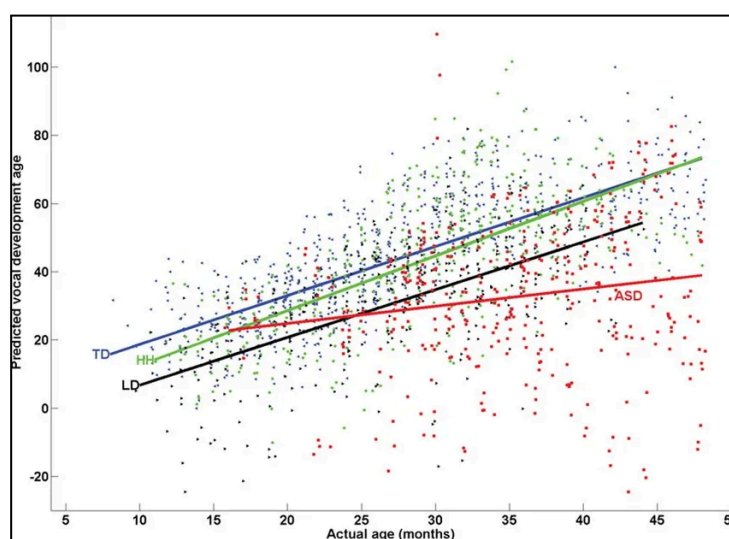


Figure 3.1 Multiple linear regression model predictions for individual observations

4. Polynomial Regression

Polynomial Regression is a supervised regression algorithm for problems with non-linear relationships. Polynomial Regression specifies the relationship between the independent variable and the dependent variable as an nth-degree polynomial, effectively expanding linear models by adding non-linear components. Model complexity is greater than linear models, with complexity rising with the order of the polynomial. As the degree of the polynomial increases, interpretability is lost, and these models become more difficult to interpret.

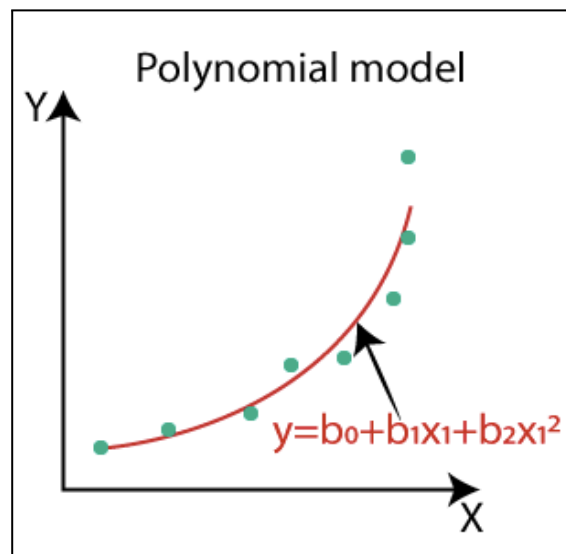


Figure 4.1 Graphical representation of the polynomial model of a regression algorithm

Training and prediction are computationally more expensive than linear models but are still within the capabilities of modern hardware. Polynomial Regression requires sufficient data points to avoid overfitting, especially with high-degree polynomials. Feature engineering includes the creation of polynomial features (e.g., x^2 , x^3) and scaling. The algorithm is sensitive to outliers and noisy data and might need smoothing or regularization. It relies on the assumption that the underlying relationship can be described by a polynomial function. Scalability is limited for large data sets or high-degree polynomials due to overfitting and computational expense. The performance measures are the same as for linear regression: MSE, RMSE, and R-squared. Regularization techniques like Ridge and Lasso avoid overfitting and improve generalization on test data. Polynomial Regression has applications in robotics for trajectory prediction, finance for non-linear risk modeling, and engineering for analyzing nonlinear systems. While it is flawed, it is still a helpful tool when the linear model fails to adequately depict the complexity of actual, multidimensional data sets.

5. Conclusion

The importance of Logistic Regression, Multilinear Regression, and Polynomial Regression extends beyond their mathematical basis; they are key components in the machine learning arsenal for converting raw information into valuable knowledge. Logistic Regression is particularly noted for being able to enhance decision-making in high-risk binary classification issues with relative simplicity and efficacy, providing the basis for a wide variety of real-time prediction models. Multilinear regression is still a very important explanatory method, providing an easy snapshot of the way multiple variables influence a quantitative response—a characteristic that continues to be of significant value across many disciplines, such as economics and operations management. Polynomial regression, since it can model non-linear trends, enables analysts to go a step further from linear models, thereby unveiling more intricate and accurate patterns in the data. Together, these algorithms illustrate the diverse approaches in the machine learning field and a compromise between adaptability and interpretability. Their continued applicability in both academic and industrial circles only underscores the necessity of grasping not only their workings but also the wherefores and whys of their application. A mastery of these core algorithms gives practitioners the toolkit to address a broad variety of analytical issues and is a starting point in the development of increasingly complex, domain-based machine learning applications.

Algorithm	Pros	Cons	Suggested Uses
Logistic Regression	Easy to interpret, fast training and prediction, good for binary outcomes	Assumes linearity in log-odds, not suitable for complex relationships	Medical diagnosis, spam detection, credit scoring
Multilinear Regression	Interpretable, models multiple features simultaneously, efficient	Requires linearity and normality assumptions, sensitive to multicollinearity	Real estate pricing, sales forecasting, economics
Polynomial Regression	Captures non-linear patterns, flexible modeling	High risk of overfitting, complex interpretation, sensitive to noisy data	Robotics, finance (non-linear modeling), engineering

6. References

Analytics Vidhya. (2021, May). Logistic Regression Algorithm | Introduction to Logistic Regression.<https://www.analyticsvidhya.com/blog/2021/05/logistic-regression-supervised-learning-algorithm-for-classification/>

Analytics Vidhya. (2021, May). Multiple Linear Regression: Definition, Examples & Applications.<https://www.analyticsvidhya.com/blog/2021/05/multiple-linear-regression-using-python-and-scikit-learn/>

Machine Learning Mastery. (2023). Capturing Curves: Advanced Modeling with Polynomial Regression.<https://machinelearningmastery.com/capturing-curves-advanced-modeling-with-polynomial-regression/>