

Chapter 1

Process of Machine Learning Methods

Hulya Kocyigit

Karamanoglu Mehmetbey University, Turkey

ABSTRACT

The field of machine learning (ML) has grown to be a prominent subject within developed businesses while aiming to implement data-driven techniques to better day-to-day business activities. The reader is introduced to the most popular learning models in this chapter. Although unsupervised learning, reinforcement learning, and semi-supervised learning are incredibly important, the authors won't go into further detail about them here. This section will go into great length about supervised learning environments. The authors propose the following as a summary of the contributions to this chapter: Emphasizing some of the earlier works of literature that tackled these problems and discussing their limitations. In order to do so, the authors propose to review the regression family (i.e., simple regression and multiple linear regression) and decision tree family (i.e., CART, ID3, C4.5, chi-squared automatic interaction detection (CHAID), bagging and boosting). Examining the function and promise of ML techniques to resolve dilemmas and present potential implementation strategies.

1. INTRODUCTION

Automated analysis strategies have been facing a unique difficulty because of developments in data gathering, storage, and processing technologies. Enormous volumes of data constantly are gathered from various sources, including stock trading, medical healthcare systems, electronic sales records, and significant scientific research. Additionally, more researchers and professionals than ever before have been striving to use automated approaches to examine their data. There has been a corresponding demand for reliable, effective, and adaptable data exploration methods as the amount and variety of data that may be accessed by these approaches expand. The ability to foresee events is an illustration of one's ability to judge others based on what they have learned. Prediction analysis has grown in popularity in a variety of fields, including spatial inference, regression analysis, and causal inference, as communities and technology have advanced.

DOI: 10.4018/978-1-6684-8386-2.ch001

Regression is the main line of research in the disciplines of machine learning and statistics. The primary goal of a regression problem is to train a learner using the available data and mapping the input to the matched output result to achieve the prediction goal. In its simplest form, regression analysis assists an organization in making more informed choices by enabling them to comprehend what their data points signify and make use of them appropriately. Given the growing importance of the business sector and the popularity of regression analysis, it is crucial for business analysts to comprehend these possibilities and be equipped to use them. Such as, if we've ever worried about how fluctuations in joblessness or inflation affect Gross Domestic Product, we need to analyze estimates relying on the relationship between covariates, and then at this point, regression analysis might be worthwhile. Note that a regression mode doesn't mean that the factors are related in a cause-and-effect manner. It is possible for two or more variables to have a strong empirical relationship, but this is not proof that the regressor variables and the answer are causally associated. The association between the regression variables and the response needs to have a foundation outside of the sample data to prove causality. The regression models will examine in the subsections of this chapter are applicable to both randomized control trial design and observational study. Regression models must have criteria that are acceptable for the data at hand, whether the data are experimental or observational, for the model to be usable.

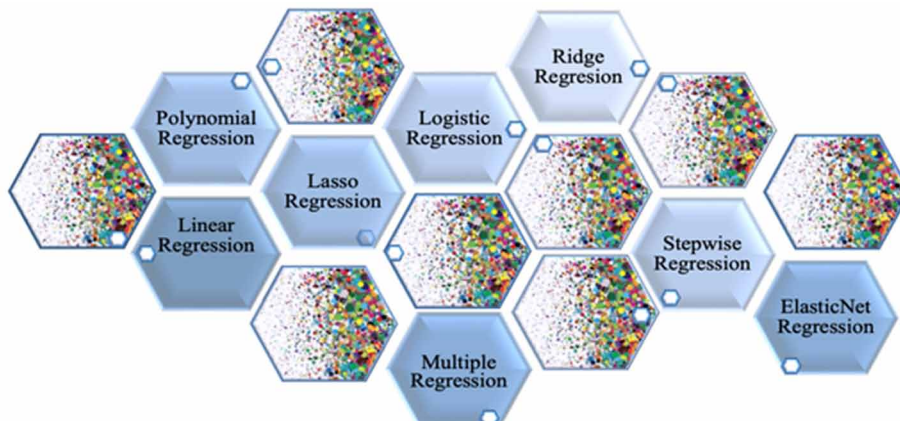
Regression model construction is an iterative procedure. Figure 1 provides an illustration of the model-building procedure.

Figure 2 illustrate types of regression methods, and the use of each method depends on the number of variables. The significance of each type varies depending on the circumstances, but fundamentally, all regression techniques examine the impact of the explanatory variables on the response variable.

Figure 1. The process of regression model



Figure 2. The process of regression model



36 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/process-of-machine-learning-methods/328295

Related Content

Application of Geographical Information System and Interactive Data Visualization in Healthcare Decision Making

Zhecheng Zhu (2016). *International Journal of Big Data and Analytics in Healthcare* (pp. 49-58).

www.irma-international.org/article/application-of-geographical-information-system-and-interactive-data-visualization-in-healthcare-decision-making/171404

Big Data Analytics in Healthcare: Applications and Challenges

Jaimin Navinchandra Undaviaand Atul Manubhai Patel (2020). *International Journal of Big Data and Analytics in Healthcare* (pp. 19-27).

www.irma-international.org/article/big-data-analytics-in-healthcare/253843

Big Data Analytics in Healthcare: Applications and Challenges

Jaimin Navinchandra Undaviaand Atul Manubhai Patel (2020). *International Journal of Big Data and Analytics in Healthcare* (pp. 19-27).

www.irma-international.org/article/big-data-analytics-in-healthcare/253843

Mapping AI's Influence in Politics Through Digital Authorship Analysis

Dimitrios Vagianosand Nikos Koutsoupas (2026). *Modern International Developments in Data Analysis and Multivariate Statistical Analysis* (pp. 303-336).

www.irma-international.org/chapter/mapping-ais-influence-in-politics-through-digital-authorship-analysis/388037

Exploratory Data Analysis

(2018). *Spatial Analysis Techniques Using MyGeoffice®* (pp. 112-136).

www.irma-international.org/chapter/exploratory-data-analysis/189720