Use of "Odds" in Bayesian Classifiers

U

Bhushan Kapoor
California State University, Fullerton, USA

Sinjini Mitra
California State University, Fullerton, USA

INTRODUCTION

The availability of massive amounts of data and the explosive growth of the field of analytics and data science has led to widespread use of machine learning in many different types of problems. Machine learning is a field of study that focuses on algorithms and techniques that learn from data or examples. Performance of machine learning algorithms and techniques improve as these are exposed to more data. According to Arthur Samuel (1959) who first coined the term machine learning, "Machine learning algorithms enable the computers to learn from data, and even improve themselves, without being explicitly programmed".

There are several categories of machine learning techniques, such as Classification, Regression, Clustering, and Association, for different types of problems. The goal of Classification techniques is to predict the output category (or label) given the input data (or predictors). As an example, loan application algorithm predicts if a loan application will be approved, or not by the financial institution. The input data or predictors in this case would be the applicant's credit history, FICO score, annual income, education, and so on. The output category (or label) that the algorithm predicts would be the different output categories, like loan is approved or loan is not approved.

Classification models belong to the category of supervised learning where the models at their development stages are provided with data/ examples on both input (predictor variables) and output category (label). After the model is built and tested, it is put in production to predict the output category (or label) given the input data (or predictors). The bayes classifiers compute prior probabilities and likelihood of input data at the development stage. At the production stage, the classifiers predict posterior probabilities for each output category (or label). In this paper we will introduce the concepts of odds, log odds, and odds ratio and reformulate Bayes' Theorem and multinomial Naïve Bayes' Classifiers in terms of odds. The reformulated bayes classifiers will compute prior odds and likelihood of input data at the development stage. At the production stage, the classifiers will predict posterior odds for each output category (or label).

This paper is divided into five main sections, (1) Introduction, (2) Concept of 'Odds' and its properties, (3) Bayes' Theorem, (4) Naïve Bayes' Classifier, and (5) Conclusion. In Section (2), concept of odds is defined in terms of probabilities. Some of its basic properties are also stated in this section. In Section (3), the authors will discuss Bayes' Theorem and reformulated Bayes' theorem using odds and probabilities. They will then present an example on loan application approval and calculate posterior probabilities and posterior odds using both forms of the Bayes' theorem. Section (4) will be devoted to Multinomial Naïve Bayes' Classifier and its reformulation using odds and probabilities. We will compare the two forms with the help of the example on loan application approval. Finally, the chapter will conclude with a summary and discussion in Section (5).

DOI: 10.4018/978-1-7998-9220-5.ch162

BACKGROUND

Like probability, 'odds' is a measure of the likelihood of an event. But these are defined differently in statistics and have different properties (Lawrence, Francis, Nathaniel & Muzaffer, 2012; Martin, 2021; Ranganathan, Aggarwal & Pramesh, 2015). The probability of an event, E (written here as P(E)) is defined as a real number that always lies between 0 and 1, and it is estimated as the number of times the event occurs over the total number of random trials or examples. When there are only two outcomes, we can use odds instead of the probability of an event. We represent the odds of an event as O(E). And it is defined as follows:

$$O\left(E\right)=\frac{P\left(E\right)}{P\left(-E\right)}=\frac{P\left(E\right)}{1-P\left(E\right)} \tag{1}$$

where -E is the complement event of E.

O(E) can assume any real number between 0 and infinity. O(E) = 1 means that the chances of event E happening or not happening are equal. When O(E) is greater than 1, it means that the chance of occurrence of event E is higher than its non-occurrence. And, when O(E) is less than 1, it means that the chance of occurrence event E is lower than its non-occurrence.

Odds of an event, E, is generally expressed in a ratio form as m:n, where m and n are positive integers (whole numbers), and O(E) = m/n. Similarly, odds of the event -E, is generally expressed in a ratio form as n:m, and O(-E) = n/m.

From the above equation (1), we have

$$P\left(E\right)=\frac{O\left(E\right)}{1+O\left(E\right)} \tag{2}$$

Given P(E) we can find O(E) by using equation (1), and similarly given O(E) we can find P(E) from equation (2). From these equations, it is obvious that

O(E)*O(-E) = 1, O(E) = 1/O(-E), and O(-E) = 1/O(E)

Some people may grasp the concept of "odds" better than probability in certain applications (Grant, 2020; Moran, 2020). For example, they get better understanding when a financial institution informs them of the odds of their loan approval rather than the probability of the loan approval.

BAYES'S THEOREM

Bayes's Theorem, named after 18th century British mathematician Rev. Thomas Bayes, is a mathematical formula that relies on incorporating prior probability distribution to generate posterior probabilities. The Bayes's Theorem is the foundation of the vast field of Bayesian Statistics, and several Machine Learning models are based on Bayesian Statistics (Tyler, Liliana & Jeong, 2018; Bolstad & Curran, 2016; Kruschke, 2014; Lee, 2012; Christensen, Johnson, Branscum & Hanson, 2011). We will explain below

Related Content