Which algorithms did you found interesting alternatives of logistic regression and why?

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Definition: | Advantages: | Disadvantages: |
| Logistic Regression | Logistic regression is a machine learning algorithm for classification. In this algorithm, the probabilities describing the possible outcomes of a single trial are modelled using a logistic function. | Logistic regression is designed for this purpose (classification) and is most useful for understanding the influence of several independent variables on a single outcome variable. | Works only when the predicted variable is binary, assumes all predictors are independent of each other and assumes data is free of missing values. |
| Naïve Bayes | Naive Bayes algorithm based on Bayes’ theorem with the assumption of independence between every pair of features. Naive Bayes classifiers work well in many real-world situations such as document classification and spam filtering. | This algorithm requires a small amount of training data to estimate the necessary parameters. Naive Bayes classifiers are extremely fast compared to more sophisticated methods. | Naive Bayes is known to be a bad estimator. |
| Stochastic Gradient Descent | Stochastic gradient descent is a simple and very efficient approach to fit linear models. It is particularly useful when the number of samples is very large. It supports different loss functions and penalties for classification. | Efficiency and ease of implementation. | Requires several hyper-parameters and it is sensitive to feature scaling. |
| K-Nearest Neighbors | Neighbors based classification is a type of lazy learning as it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the k nearest neighbors of each point. | This algorithm is simple to implement, robust to noisy training data, and effective if training data is large. | Need to determine the value of K and the computation cost is high as it needs to compute the distance of each instance to all the training samples. |
| Decision Tree | Given a data of attributes together with its classes, a decision tree produces a sequence of rules that can be used to classify the data. | Decision Tree is simple to understand and visualize, requires little data preparation, and can handle both numerical and categorical data | Decision tree can create complex trees that do not generalize well, and decision trees can be unstable because small variations in the data might result in a completely different tree being generated. |
| Random Forest | Random forest classifier is a meta-estimator that fits several decision trees on various sub-samples of datasets and uses average to improve the predictive accuracy of the model and controls over-fitting. The sub-sample size is always the same as the original input sample size, but the samples are drawn with replacement. | Reduction in over-fitting and random forest classifier is more accurate than decision trees in most cases. | Slow real time prediction, difficult to implement, and complex algorithm. |
| Support Vector Machine | Support vector machine is a representation of the training data as points in space separated into categories by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. | Effective in high dimensional spaces and uses a subset of training points in the decision function so it is also memory efficient. | The algorithm does not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation. |

Main disadvantage of the logistic regression is too many assumptions and limitations on the type of predictor that is used. Besides that, this algorithm works just fine for me. The above table reflects pros and cons of each of the 7 Types of Classification Algorithms, presented on https://analyticsindiamag.com/7-types-classification-algorithms/. Based on that, and my current situation regarding computing power and time availability, my choice goes with Support Vector Machine, Stochastic Gradient Descent and Naïve Bayes. My reasoning is simple – my computing abilities at the moment are extremely limited so if I am required to use them on my machine, I will choose the most efficient ones regarding performance. Nevertheless, using cloud software as Jupyter Notebook, it is not a guarantee that the task will be complete in a timely manner. For instance, last week I was trying few algorithms, discovered on Stack Overflow, and the Kernel got busy indefinitely. My attempts to interrupt the process were futile, so after few hours of waiting (actually doing my job while waiting) I was left with no choice, but to restart the server. It all depends on what resources you got, or what type of access you have. I tend to have some sympathies directed to K-Nearest Neighbors algorithm since it handles large and noisy data. KNN can be used for both classification and regression problems. The K-Nearest Neighbors algorithm uses ‘feature similarity’ to predict the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set. Another alternative of logistic regression algorithm is Random Forest. This one look complicated enough and possess enough skills to assassinate my CPU. Random Forest is a popular machine learning model that is commonly used for classification tasks. In addition to classification, Random Forests can also be used for regression tasks. A Random Forest’s nonlinear nature can give it a leg up over linear algorithms, making it a great option. However, it is important to know your data and keep in mind that a Random Forest can’t extrapolate. It can only make a prediction that is an average of previously observed labels. In this sense it is very similar to K-Nearest Neighbors. In other words, in a regression problem, the range of predictions a Random Forest can make is bound by the highest and lowest labels in the training data. In conclusion, no matter of what my technical capabilities or personal opinion are, each of those 7 algorithms exists for a reason, and can be suited to certain task. It would be best to test all or most of the models and/or consider using them in an ensemble of models.