Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 18, No. 2, 2026

A Review of the Principles and Applications of Mainstream Classification Algorithms: From Logistic Regression to Neural Networks

Xinyang Zhao *

School of Information Engineering, Minzu University of China, Beijing, 100081, China
*** Corresponding author Email:** 2583057842@qq.com

Abstract: This article illustrates four mainstream classification algorithms: logistic regression systematically, K-nearest neighbors (KNN), decision trees, and neural networks. Analyzed the mathematical principles, core calculation formulas, and basic achieving methods of each of these four kinds of algorithms; From the perspective of practical application scenes, analyze the applicability and obvious limitations of these algorithms when facing different data. This study provides a systematic proposal for the theoretical understanding and practical application of classification models in engineering and daily life.

Keywords: Classification Algorithm; Logistic Regression; K-nearest Neighbor; Decision Tree; Neural Networks.

1. Chapter 1: Introduction

Recently, the mainstream of artificial intelligence is data intelligence. To some extent, there is no data, there is no intelligence, and there is no data, there is no explanation [1]. Algorithms can adjust and correct itself through feedback circuits during execution process, ensuring continuous optimization of classification standard and logic. [2]. This article analyzes these four mainstream algorithms from the aspects of algorithm principles, core operating mechanisms, and practical application scenes in a profound way. Combined with data features, it searches the practical applications of these algorithms, hoping to provide classification algorithm selecting guidance and reference for researchers and engineering practitioners in related fields.

"Classification" is the essential method for people to recognize the objective world. When facing the objective world experienced, people usually pay attention to two aspects: "similarity" and "difference"[3]. Classification is a significant research field in data mining, which involves training a classifier in a group of samples with known class labels to classify an unknown sample. The classification process of a classification algorithm is to establish a classification model to describe a predetermined dataset or concept set, and construct the model by analyzing database tuples described by attributes[4].Data mining is the process of obtaining valuable knowledge patterns from large-scale data, mainly using data mining algorithms to realize[5] .In general, the purpose of designing classification algorithms is to achieve a so-called "classifier" (program) [6] .

Building a classification model can be divided into the following steps: data preprocessing work, feature engineering, conducting model training, completing hyperparameter optimization and performance evaluation, and testing the practical application ability of the model on existing test sets. By following this process, I firmly believe it can provide a reliable and operational framework for the practical application of various classification algorithms in daily life.

2. Chapter 2: Logistic Regression

2.1. Basic Principles of the Model

Logistic regression is a widely used probabilistic linear model in classification. Although the word "regression" may leave a vague impression on us, it indeed belongs to classification algorithms. When using logistic regression algorithm for classification, the final goal is to obtain the best coefficients[7].Logistic regression is suitable for classification problems, but when coping with some problems, in order to make the algorithm more fitting, it is necessary to construct a hypothesis function that is as suitable as possible for all data[8].Logistic regression is a general linear regression model that shares many similarities with multiple linear regression. Unlike multiple linear regression, it incorporates the logistic (sigmoid) function, making it widely used for binary classification problems to estimate the probability of an event occurring[9]

2.2. Mathematical Formulation and Training Mechanism

Let the input feature vector be $\mathbf{x}=(x_1,x_2,\ldots,x_n)^{\top}$, The model parameters include the weight vector $\mathbf{w}=(w_1,w_2,\ldots,w_n)^{\top}$ and the bias term b, Then, the predicted probability of the positive class (e.g.(y = 1)) is:

$$P(y=1|\mathbf{x})=\sigma(\mathbf{w}^{\top}\mathbf{x}+b)=\frac{1}{1+e^{-(\mathbf{w}^{\top}\mathbf{x}+b)}}$$

where σ (·) is the Sigmoid function. The classification decision rule is typically given by: If $P(y=1|\mathrm{x})\geq 0.5$, then predict the positive class. otherwise, predict the negative class.

The model is trained by minimizing the cross-entropy loss function:

$$\mathrm{L}(\boldsymbol{w},b)=-\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\hat{y}^{(i)}+\left(1-y^{(i)}\right)\log\left(1-\hat{y}^{(i)}\right)\right]$$

Where $\hat{y}^{(i)}=P\left(y=1|\mathbf{x}^{(i)}\right)$. By using optimization algorithms such as gradient descent, the parameters are iteratively updated to ensure convergence of the loss function.

2.3. Practical Application

The logistic regression model itself not only has high prediction accuracy, but also is easy to understand and meets regulatory requirements for explainability. A significant step in the practical application of logistic regression algorithm is to establish feature engineering, which aims to extract features from raw data to the maximum extent possible for using by algorithms and models[10].Logistic regression can only learn to construct linear decision boundaries; consequently, when confronted with data distributions exhibiting highly nonlinear characteristics, the overall performance of the model becomes limited.

Logistic regression is widely used in finance for credit scoring and risk prediction; in healthcare for modeling health risk assessment; and in the internet industry for forecasting trending traffic. Owing to its reliability, interpretability, and transparency, logistic regression serves as an excellent baseline model for classification tasks, providing a benchmark against which the performance of more complex algorithms can be evaluated.

3. Chapter 3: K-Nearest Neighbors Algorithm

3.1. Basic Principles of the Model

The K-Nearest Neighbors (KNN) classification algorithm is a theoretically well-established method and one of the simplest machine learning algorithms. The core idea of this approach is: in the feature space, if the majority of the K nearest neighbors (i.e., the most proximate samples in the feature space) of a given sample belong to a certain class, then the sample is also assigned to that class[4]. In real-life scenarios, most problems consist of multiple factors that interact with—and often conflict with—each other. People frequently encounter situations where they must weigh multiple factors simultaneously to arrive at a final decision[11].

3.2. Mathematical Formulation and Implementation Mechanism

Let the sample to be predicted be $\mathbf{x}$, Let the i-th sample in the training set be $\mathbf{x}^{(i)}$, Similarity between them is typically measured using Euclidean dista*nce:*

$$d(\mathbf{x},\mathbf{x}^{(i)})=\sqrt{\sum_{j=1}^{n}(x_j-x_j^{(i)})^2}$$

Manhattan distance, cosine similarity, etc. can also be used depending on the features of the data. The algorithm first calculates the distance between x and all the training samples, selects the K neighbors with the smallest distance, and then ensures the prediction category through majority voting:

$$\hat{y}=arg\max_{c\in Y}\sum_{i\in N_K(\mathbf{x})}I(y^{(i)}=c)$$

Among them, $N_K(\mathrm{x})$ illustrates the K nearest neighbor sets of x, and $I(\cdot)$ is the indicator function. To increase accuracy,we can use distance weighted voting to increase neighbors' contribution.

3.3. Practical Application

The KNN algorithm model is easy to understand, and it performs very stably on datasets with small sample and clear boundaries between categories. But the computing cost is relatively high and not suitable for scenes that need real-time processing or large amounts of data; Being severely influenced by featuring scale, it must standardize the traits before using.

The KNN algorithm is usually used in tasks with non-linear data and relatively clear local structures. For instance, in the task of recognizing handwritten digit; In recommendation systems, it can also be used to match similarity between users or items. Additionally, the KNN algorithm is often used in circumstances such as prototype verifying and teaching demonstrations.

4. Chapter 4: Decision Tree

4.1. Basic Principles of the Model

A decision tree is a kind of tree structure classification model. At each internal node of the tee structure, selects the optimal feature, divides the current sample set into various subsets, and finally, each path from the root node of the tree to each leaf node all corresponds to a specific classification judging rule. Based on the clear and direct feature of decision tree model's classification method, making it easy to understand and explain. At the same time, as a very fast learning and prediction algorithm, it can provide relatively high text classification efficiency and is suitable for handling large-scale text data at classification scenes [12].

4.2. Classification Criteria and Mathematical Foundations

By using decision tree classification algorithm for classification, the appearing classification rules are relatively easy to understand, and because each branch in the decision tree corresponds to different classification rules, thus during the final classification process it can generate a set of rules easier to understand. In the process of using decision tree classification algorithm to classify data in data mining, compared with other classification methods, it has faster velocity and higher efficiency[13].

A decision tree is a multi-level tree, which has a root node (sample set), intermediate nodes (decision process), and leaf nodes (decision results) that progress layer by layer[14].The commonly used feature selection indicators have information gain, information gain rate, and Gini impurity these three kinds. The information gain is the core indicator of the ID3 algorithm, which is based on Shannon entropy and used to measure the degree of reduction in overall uncertainty of the dataset after feature divding:

$$Gain(D,A)=H(D)-\sum_{v=1}^{V}\frac{D_v}{D}H(D_v)$$

Among them, $H(D)=-\sum_{k=1}^{K}p_k\log_2 p_k$ is the entropy of dataset D, D_v is the subset divided according to feature A.

Gini Impurity (CART algorithm) judges the probability of randomly selected two samples have different categories:

When dividing $Gini(D)=1-\sum_{k=1}^{K} p_k^2$, select the features and segmentation points that minimize the weighted Gini index. The growth of trees usually pre-prune (limiting depth and minimum sample size) or post pruning (building trees first and then simplifying) to prevent overfitting.

4.3. Practical Application

The current research hot spot of decision tree algorithm is to solve rule replication problems and update speed problems while ensuring algorithm performance, and to get a better compromise between important indicators such as algorithm throughput, internal memory consumption, and update velocity[15].Decision trees are easy to be overfitting and are overly sensitive to small changes in data without proper processing. Therefore, a single decision tree is not used in most of time.

Decision trees can make clear judgments in credit examination work in the financial field; In the medical field, it can also evaluate health risks; In customer classification work, it can quickly and accurately find customer's features. Some algorithms of decision trees have become the core models of many big data analyses and are important basic algorithms for modern ensemble learning.

5. Chapter 5: Neural Networks

5.1. Basic Principles of the Model

Neural networks are a type of nonlinear model inspired by the basic principles of biological nerve systems. They integrate various parameterized functions to realize accurate data reflection. In classification tasks, its form is Multi Layer Perceptron (MLP), which consists of an input layer, one or more hidden layers, and an output layer. Each layer has many neurons, which are computed layer by layer to get the results. Artificial Neural Network (ANN) is a new generation of artificial intelligence information technology with significant application value. It can effectively learn and handle complex tasks and has been widely used in various fields such as image handling, voice interaction, and autonomous driving [16].

5.2. Mathematical Formulation and Implementation Mechanism

Let the input to the l -th layer be $\mathbf{a}^{(l-1)}$, the weight matrix be $\mathbf{W}^{(l)}$, the bias vector be $\mathbf{b}^{(l)}$, then the linear transformation of this layer is:

$$\mathbf{z}^{(l)}=\mathbf{W}^{(l)}\mathbf{a}^{(l-1)}+\mathbf{b}^{(l)}$$

Introducing nonlinearity through activating functions $\phi(\cdot)$such as ReLU, Sigmoid, Tanh):

$$\mathbf{a}^{(l)}=\phi\left(\mathbf{z}^{(l)}\right)$$

For binary classification tasks, the output layer usually uses the Sigmoid function; For multi classification, the Softmax function generates probability distributions. This model depends on the chain method to calculate each parameter, and then updates the weight data to make the prediction error extremely small.

5.3. Practical Application

In the direction of computer vision, neural networks are used for multi scene applications such as image classification; In the field of natural language handling, recurrent neural networks (RNNs) and transformers are widely used in sentiment analysis and email classification handling; In the industrial sector,it can be used for defect classification checking; In biology, it can predict the function of gene sequences greatly. Because of its accurate advantage, it is also the best approaches for classification problems.

6. Chapter 6: Comprehensive Comparison and Selection Suggestions

These four algorithms have absolutely different modeling ways. Logistic regression is a parameterized linear model, K-nearest neighbors is a non parametric lazy learning method, decision trees are divided recursively, and neural networks are changed through multiple layers of nonlinear transformations.

If emphasizing explainability and compliance, logistic regression or decision tree should be take into consideration first; If the data scale is small, the dimension is low, and the distribution is non-linear, K-nearest neighbors should be prioritized; If pursuing high precision and having sufficient data and extraordinary computing ability, neural networks are usually the first choice; If you need a balance between performance and transparency, you can consider integration methods.

There is no absolute "best" algorithm, only have a more suitable algorithm. In practical applications, start with simplicity and then complexity. Use logistic regression and decision trees to build a basic model, and then based on effect to switch to more complex methods. Only by combining cross validation, parameter tuning, and actual business situations can make the classification model be effectively used and applied.

7. Chapter 7: Conclusion

This article illustrates the basic principles, mathematical mechanisms, and application scenarios of four types of algorithms, analyzes their features and limitations in terms of computing speed, model structure, data requirements, and robustness. Through further comprehensive comparison we can perceive that selecting algorithm should not only pursue the best performance indicator, but also take into account the actual task situation, such as the number of data, the high and low feature dimensions, whether need real-time processing, whether can understand the model logic, and whether can understand the existing resource conditions. After comprehensive consideration, then make the best selection. Choosing proper classification algorithms will assist stimulate the rational application and continuous innovation of classification technology in different fields. The practical method proposed in this article, from simple to complex, can provide clear and useful references for researchers and engineers when choosing appropriate models.

References

[1] Li, G. J. (2024). Big data and computational models. Big Data, (1), 9–16. (in Chinese)

[2] Zhang, M. J., & Guo, X. A. (2025). The "order" of technology: Classification logic of intelligent algorithms and "substitution by generalization". Modern Publishing, (7), 51–63. (in Chinese)

[3] Wu, T. (2000). Classification and bifurcation: An exploration of the self-organizing origins of knowledge and scientific knowledge — Taking the evolution of classification and the shift of cultural perspective on classification in biology as an example. Journal of Dialectics of Nature, (6), 25–33. (in Chinese)

[4] Meng, L., & Liu, J. P. (2022). Research on classification algorithms in data mining. China Education Network, (7), 43–45. (in Chinese)

[5] Xu, Y., & Qu, D. Q. (2024). Design and implementation of computer network intrusion detection system based on decision tree classification algorithm. Information Recording Materials, 25(6), 137–139. (in Chinese)

[6] Li, X. M. (2020). Algorithms in classification problems. Chinese Information Technology Education, (23), 24–28. (in Chinese)

[7] Deng, J. Y. (2019). Research and application of multinomial logistic regression algorithm in the context of big data [Master's thesis, Wuhan Institute of Technology]. (in Chinese)

[8] Liao, L., Chen, Y. Y., Zeng, G. F., & Liu, P. Q. (2025). Logistic regression classification method based on granular computing. Computer Simulation, 42(2), 382–388. (in Chinese)

[9] Xia, M. (2024). Research and application of fusion algorithm of decision tree and logistic regression model in binary classification problems [Master's thesis, University of International Business and Economics]. (in Chinese)

[10] Tang, T., & Li, M. (2022). Analysis and research on logistic regression algorithm in the context of big data. Electronic Technology & Software Engineering, (17), 177–181. (in Chinese)

[11] Han, L. R. (2019). Research on k-nearest neighbor and reverse k-nearest neighbor query based on comprehensive evaluation in obstacle space [Master's thesis, Yanshan University]. (in Chinese)

[12] Liu, X. M., Li, C. Z. X., Wu, S. C., Zhang, Y. C., Bai, H. Y., Cheng, Z. H., Chen, Z., Li, Y. F., Lan, Y., & Shen, C. (2024). Review on text classification algorithms and their application scenarios. Chinese Journal of Computers, 47(6), 1244–1287. (in Chinese)

[13] Cheng, Y. F. (2021). Overview of data classification algorithms in data mining. Digital Communication World, (2), 136–160. (in Chinese)

[14] Wang, Y. L. (2022). Research on machine learning algorithm classification based on artificial intelligence. Computer Fan (Popular Edition), (7), 22–24. (in Chinese)

[15] Lü, G. F., Tan, J., Qiao, G. J., & Yan, J. L. (2022). Decision tree algorithm for message classification. Journal of National University of Defense Technology, 44(3), 184–193. (in Chinese)

[16] Jiang, S. Q. (2025). Research on supervised learning algorithm of deep spiking neural network and its application [Doctoral dissertation, Xidian University]. (in Chinese)