Feature-Enhanced Data Classification Using B-center Function

Enyu Li 1,*,†, Leyu Lyu 2,† and Wenfei Mao 3,†

1School of Business, Jilin University, Changchun, Jilin Province, 130000, China;
2School of Information Engineering, China Jiliang University, Hangzhou, Zhejiang Province, 310016, China;
3School of Statistics and Big Data, Henan University of economics and law, Zhengzhou, Henan Province, 450000, China;
*Corresponding author e-mail: liey3418@mails.jlu.edu.cn
†These authors contributed equally.

Abstract. Linear-kernel Support Vector Machine (SVM) classification algorithm needs to constantly calculate the similarity between the data to be tested and the sample during classification. If samples in the dataset are linearly indivisible, the accuracy of the algorithm will get worse. The feature weight calculation based on category information in Multi-layer Perceptron (MLP) algorithm is not accurate enough to express the relationship between features and categories, i.e. it cannot distinguish categories precisely when the features with the same category frequency. In order to improve the performance of MLP and linear-kernel SVM algorithms in classification, an improved function: the Base function, which is based on clustering, is proposed in this paper. Before running the classification algorithm, a series of possible centers is predicted, after which the Base function is used for data clustering to increase the dimension of data, and then samples with more features are passed to classification models. In the high-position Gaussian data space, the Base function can obviously improve the spatial distribution of samples. Experimental comparison in the thesis also proved that the Base function can improve the performance of MLP or SVM in classification effectively. The proposed feature-enhanced MLP is superior to the baseline with 32.6% accuracy promotion on the Gaussian dataset.

1. Introduction

In the field of computer vision, categorization algorithms are mainly used for image recognition, such as fingerprint recognition and categorization [1] in criminal cases, and individual detection and counting of fruits and vegetables in agriculture [2]. In addition, concerning the field of natural language processing, the main application of categorization algorithms is text recognition, such as the categorization of Internet forum [3] opinions in multiple languages using the sentiment analysis method. Besides, as for the field of language recognition, categorization algorithms are mainly used in speech control and interaction of micro-robots[4] as well as special chips[5] for speech signal processing.

The raw space of the data provided is finite-dimensional. Therefore, the classification performance can be improved by increasing the sample feature dimension in the low-dimensional data. At present,
the popular classification algorithms are Support Vector Machine (SVM), Multi-layer perceptron (MLP) and so on. SVM is a substantially linear categorization algorithm whose training model aims to find the separating hyperplane vector $w$ between two different categories so as to maximize the separation distance. SVM also includes the kernel trick, which makes it essentially a nonlinear categorizer. The learning strategy of SVM is interval maximization, which can be formalized as a problem of solving convex quadratic programming, that is, the problem of minimizing a regularized hinge loss function. In fact, the learning algorithm of SVM is the optimization algorithm for solving convex quadratic programming. MLP is a forward structured artificial neural network (ANN) that maps a set of input vectors to a set of output vectors. MLP is a typical feed-forward network based on back-propagation learning, and its information processing direction is from the input layer to each hidden layer to the output layer, thus realizing the process mode of layer by layer. The hidden layer implements a nonlinear mapping to the input space, while the output layer implements a linear categorization, in which the nonlinear mapping approach and the linear discriminant function can be learned simultaneously.

Kernel function in SVM and activation function in MLP is used to deal with linear non-divisible data. Feature enhancement-based image recognition techniques play a crucial role in various fields. Bernard[6], Olivier and van Assen[7] M have optimized image reconstruction and image quality for imaging in the field of cardiac radiology in clinical medicine, accomplishing tasks to assist in segmentation, quantification and reporting. Guo[8], H et al. demonstrated a microwave (MW) cavity interferometric enhancement method for imaging nano defects on metal waveguide surfaces, which can be used for online evaluation and screening of chips. In NLP-oriented algorithms, feature enhancement techniques are specifically embodied in text feature processing. Wei Jason and K-Zou[9] used EDA's four data feature enhancements to effectively improve model performance for text categorization, thereby preventing model overfitting and improving model generalization capabilities. In algorithms for language recognition, n-gram feature processing work can effectively add important text features to model training and enhance model evaluation metrics[10]. Overall, feature enhancement techniques can make recognition algorithms more accurate and human-computer interaction more concise.

The selection of kernel function and feature enhancement processing before substitution into the MLP model will directly affect the accuracy of data classification. In order to overcome the problem of poor model classification performance in the face of fewer key features of data and improve the accuracy of data classification, this paper combines the characteristics of SVM and MLP algorithm and multi-dimensional Gaussian dataset, introduces the b-center function to optimize the SVM and MLP classification algorithm.

This study proposes a new feature enhancement method with b-center function. The newly introduced b-center function owns the ability to map low-dimensional data into multi-dimensional geometric space. It can remove the information redundancy among variables and reduce the correlation across features through the method of feature visualization, and then train the classifier according to these features. In this paper, b-center function is introduced as a new kernel function to combine with the SVM algorithm to improve the classification ability of the model. At the same time, the b-center function is used for features augmentation of the data to improve the classification accuracy of the MLP algorithm. Compared with the SVM algorithm with linear kernel function and MLP algorithm without feature augmentation, this paper finds that the SVM algorithm can achieve the highest prediction accuracy of 90.2% after b-center function is adjusted, and the accuracy of the MLP algorithm can reach the highest of 88.0%. Compared with the traditional SVM and the traditional MLP classification model, their performances are improved to different degrees.
2. Method

2.1 Overall workflow of our method
The classification task is to provide the object class using the proposed method. In this task, the data often obeys specific distribution, which can help classify the data effectively. This research aims to propose a feature enhancement method to improve the performance of classification methods. The method can be separated into 3 modules, including calculating the basis vector, applying weighted linear combination, data classification. The specific flow chart is shown below.

![Proposed feature enhancement method](image)

Figure 1. Classification with the proposed b-center function flow chart.

The process of using the b-center function to expand the feature to multiple dimensions is as follows: Firstly, the data needs to be clustered to find the basis vector. Secondly, multiple basis vectors are expanded into a vector space and then use the vector space to classify the data in two or more categories. Eventually, by doing previous steps, the obtained results would be found by classifying in a multi-dimensional space are more ideal.

2.2 B-center function
In this part, the purpose is to compare which classification result is more ideal in a two-dimensional plane and a multi-dimensional space. In order to achieve this goal, up-dimensional processing is needed to go through to transfer the b-center function from two dimensions to multi-dimensional space. In other words, that is, extracting multiple feature values are changed from extracting two feature values. Besides, parameter $r$ means the radius of the circle and the parameter $r$ needs to be lower for dense datasets. This is because that dense data can be classified easily in a higher-dimensional space. By doing previous steps, classification results can be compared in a two-dimensional plane and multi-dimensional space. The final results by classifying in a multi-dimensional space are more ideal. The function of b-center function is clustering points that are close together into one category. The b-center function is used for getting centers and weights. The following steps are how to get centers and weights.

2.2.1 Clustering initialization
Firstly, the sample needs to be traversed to find points. At the same time, the points found require to be ensured that are not in any center circle. Besides, in order to avoid isolated nodes, there should have multiple points by using the min_Neighbors parameter. The usage of the min_Neighbors parameter is to ensure that there must be enough points in a circle to avoid isolated points. In this way, the appearance of poor classification results can be avoided. Secondly, for points in the circle, the average of the X-axis and Y-axis coordinates should be calculated respectively. In order to ensure that the sum of squared distances from center to other points is the smallest. Thirdly, because the points that outside the circle will move inside the circle, the previous steps should be repeated until breaking the cycle. Then the local optima will be reached.
2.2.2 Ways to get weights

From the previous steps, the centers of the clustering have been got. This section aims to calculate the weights for the basis vector to combine them linearly. The obtained basis function is embedded using equation (1) as shown below.

\[
F(x) = \sum_{j=0}^{k-1} weights \times b(x,center[j]) \tag{1}
\]

Firstly, \(F(x)\) will be a linear combination of \(b(x,center[j])\) and \(F(x)\) is the enhanced feature, \(b(x,center[j])\) and the weights is obtained in this section. \(K\) is the number of clustered center, which is a data-driven parameter and fit the input data. Note that the function of the \(b(x,center[j])\) is to expand the number of features to center.

Secondly, From the \(b(x,center[j])\) function, each \(x\) value has a \(b(x,center[j])\) function value as the corresponding and each circle with center as the center can cover multiple \(x\) values. Besides, the Euclidean distance needs to be added between each center-centered circle point and center as the value of the \(b\)-center function corresponding to each other.

Thirdly, the \(b\)-center function needs to be stored corresponding to each center into a new matrix, as the new input value in \(F(x)\), you can directly use the logistic regression method to update the weight parameters. Note that it can be understood that the activation function is the sigmoid function, and the loss function is the cross-entropy function of the shallow neural network back propagation optimization weights.

Finally, through optimization, the final optimized value of weights parameters can be obtained, and the number of them is consistent with the number of center, and then can be put into the \(F(x)\) function to predict.

2.3 Feature linear combination

The enhanced features are obtained via Equation (1). Then, the feature is used to classify the data category using MLP [11], which can be expressed as MLP(\(F(x)\)). Then these input data will be classified into corresponding labels. Followed by the MLP, a softmax layer is added to produce the labels. The formulation of the softmax function is shown as follows.

\[
\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad j=1,2,\ldots,k \tag{2}
\]

In the two-dimensional space, when the output results are greater than 0.5, these output data will be classified in category 1. Conversely, when the output results are less than 0.5, these output data will be classified in category 0.

2.4 Implementation details

In the part of optimizing centers, the max iteration is set to 50, and the parameter min _Neighbors is 1/200 of sample size. In classification, a 3-layer neural network is built. Its first and second layer has 64 cells and chooses ReLU as its activation function. The softmax function is chosen as the activation function in the third layers to solve the multiple-class problem. Moreover, the cross-entropy cost function is used as our loss function. For SGD, the learning rate is 0.01. Decay is set to 1e-6, and momentum is 0.9. In data generation, the size of the dataset is 1500. The training and test dataset are split up with 2:1 and the number of each subset is 1000 and 500, respectively.

The code is implemented by Python 3.8 with Keras-2.4.3 and Tensorflow 2.4 frameworks and tests on the PC with Intel I5-8300H@2.30GHz and Nvidia GTX 1070 Max-Q with 16G RAM.
3. Experiments

3.1 Data description
The dataset is based on the Gaussian mixture model. It contains several clusters(c_i), and data \( \sim N(c_i, Var) \) (Var is chosen randomly). At first, parameters are set such as nSamples (the size of the dataset), nDim (the dimension of each sample), nClasses (the number of classes), nClusters (the number of clusters). Then n clusters are chosen randomly and generate samples with Gaussian distribution function for each cluster.

3.2 Experimental results
Accuracy is used to evaluate the performance of the model, which is successfully predicted samples / all samples, which can be expressed as,

\[
    Acc = \frac{Num_{correct}}{Num_{all}} \tag{3}
\]

To evaluate the improvement with b-center function, comparing the accuracy of using b-center function or not is necessary. As Table 1 shows, b-center function make difference in classification.

| Index | MLP Classification | MLP Classification with b-center Function |
|-------|--------------------|------------------------------------------|
| 1     | 0.536              | 0.862                                    |
| 2     | 0.591              | 0.892                                    |
| 3     | 0.554              | 0.856                                    |
| 4     | 0.572              | 0.902                                    |
| 5     | 0.528              | 0.899                                    |
| Avg.  | 0.556              | 0.882                                    |

3.3 Feature enhancement with various classification methods
To show the performance of b-center function with different classification model, SVM is also used with the linear kernel to classify the same dataset. As Table 2 shows, b-center is effective for another model.

3.4 Visualization
As Figure 2 illustrates, different colours points are used to present the result of the prediction, and black point means failed prediction. In addition, a large circle describe centers that found, and its radius is parameter r.

| Index | SVM Classification | SVM Classification with b-center Function |
|-------|--------------------|------------------------------------------|
| 1     | 0.394              | 0.840                                    |
| 2     | 0.538              | 0.880                                    |
| 3     | 0.418              | 0.862                                    |
| 4     | 0.472              | 0.870                                    |
| 5     | 0.356              | 0.834                                    |
| Avg.  | 0.436              | 0.857                                    |
Figure 2: Visualization of the classification results. The method can classify the data with multiple classes. Here, nClasses = 3 is taken as an example.

4. Conclusion
This paper proposes that the accuracy of recognition and the accuracy of point classification can be improved when using the b-center function. There are two main components in this research: first, the data needs to be clustered to find the basis vector, then multiple basis vectors are expanded into a vector space to classify the data in two or more categories. Experiments on the Gaussian mixture model dataset show this method is capable of handling the classification task with high precision via the proposed b-center function. The application of the method is using for text recognition, language recognition and so on.

References
[1] Win K N, Li K, Chen J, et al. Fingerprint classification and identification algorithms for criminal investigation: A survey[J]. Future Generation Computer Systems, 2019, 110.
[2] Bargoti S, Underwood J P. Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards[J]. Journal of Field Robotics, 2017, 34(6):1039-1060.
[3] Abbasi A, Chen H, Salem A. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums[M]. ACM, 2008.
[4] Sanchez-Rada, J, Fernando, et al. Enhancing deep learning sentiment analysis with ensemble techniques in social applications[J]. Expert Systems with Application, 2017.
[5] Kandagal A P, Udayashankara V. Automatic bimodal audiovisual speech recognition: A review[C]// 2014 International Conference on Contemporary Computing and Informatics (IC3I). IEEE, 2015.
[6] Bernard O, Lalande A, Zotti C, et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved?[J]. IEEE Transactions on Medical Imaging, 2018:1-1.
[7] Assen M V, Muscogiuri G, Caruso D, et al. Artificial intelligence in cardiac radiology[J]. La Radiologia Medica, 2020.
[8] Guo H, X Li, Zhu Q, et al. Imaging nano-defect of metal waveguides using the microwave cavity interference enhancement method[J]. Nanotechnology, 2020, 31(45).
[9] Wei J, Zou K. EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks[J]. 2019.
[10] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond Empirical Risk Minimization[J]. 2017.
[11] Rosenblatt, Frank. x. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC, 1961