

Feature selected based on PCA and optimized LMC

Ke Xi¹, and Cheng Cai^{1,*}

¹School of Electronic Information Engineering, Shanghai Dianji University, Shanghai, China

Keyword: PCA, LMC, Optimized LMC, Feature selection.

Abstract. In this article, we propose an optimization algorithm for the original LMC [1] (Large Margin Classifier). We use PCA [2] (Principal Component Analysis) to reduce the dimensionality of the images, and then put the data after dimensionality reduction into the optimized LMC for the feature selection [3]. We will get several features with the greatest distinction. We use these features to classify images. Finally, the experiment shows that the accuracy of the optimized LMC under the same dimensions is higher than that of the original LMC, and in many cases, the accuracy of the optimized LMC after taking 6 feature vectors has exceeded the highest accuracy of the original LMC.

1 Introduction

With the improvement of computer computing power, machine learning has made great breakthroughs in recent years. Based on machine learning, there are many classification models such as Faster-rcnn, Yolo, SSD, and Resnet. There are also many classification algorithms such as SVM [4] (Support Vector Machine), DT [5] (Decision Tree), LR [6](Logistic Regression) and many other algorithms. In this paper, the algorithm we use is a simplified version of SVM named LMC. We have optimized the original LMC based on PCA to get the feature selection. The original LMC only circulates the feature vector according to the energy value from large to small after PCA dimensionality reduction. Although the energy of the first feature vector is the largest, its distinguishability is not necessarily the best. Therefore, we have taken this point and then optimized it.

2 Principal component analysis

PCA is the most widely used algorithm of data dimensionality reduction in this world. The main idea of PCA is to map n -dimensional features to k -dimensional features, and the k -dimensional features are the new orthogonal features called principal component. In fact, by calculating the covariance matrix of the data matrix, we will get the eigenvector, and

* Corresponding author: caic@sdju.edu.cn

then we select the matrix composed of the eigenvectors with the largest eigenvalue. In this way, data dimensionality can be achieved.

The main advantage of PCA is to maintain the original data information to the greatest extent.

3 Large margin classifier

LMC uses the concept of convex hull, but the condition of convex hull is too harsh. So in this paper, we use the concept of affine hull. The sample points are described as subspaces X which can be written as:

$$X = \sum_{i=1}^n \alpha_i x_i (\alpha_i \geq 0, \sum_{i=1}^n \alpha_i = 1) \tag{1}$$

while we use the concept of affine hull, the formula can be written as:

$$X = \sum_{i=1}^n \alpha_i x_i (\sum_{i=1}^n \alpha_i = 1) \tag{2}$$

This change relaxes the requirements on α , and we expect it to be linearly related. If the subspaces overlap, the meaning of our classification will not exist. So we assume a positive example X_+ and a negative example X_- under the premise of linear separability. We can write them separately as:

$$X_+ = U_+ v_+ + \mu_+ \quad X_- = U_- v_- + \mu_- \tag{3}$$

In this formula, U is the feature vector, v is the corresponding coefficient in the feature vector projection, μ is the mean. Their subscripts indicate whether they are positive example or negative example.

The idea of LMC is to find the two closest points between two subspaces, and then find the normal vector of the distance between these two points so that we can separate the two samples. We can write the distance d_{min} between two subspaces as

$$d_{min} = \|(U_+ v_+ + \mu_+) - (U_- v_- + \mu_-)\|^2 \tag{4}$$

$$U = (U_+ \quad -U_-) \quad v = \begin{pmatrix} v_+ \\ v_- \end{pmatrix} \quad \mu = \mu_+ - \mu_- \tag{5}$$

With the formula 5, the formula can be written as

$$d_{min} = \|Uv - \mu\|^2 \tag{6}$$

We can get v by deriving v in formula 6.

$$v = (U^T U)^{-1} U^T \mu \tag{7}$$

The split plane w and offset b in the decision function of $f(x) = \langle w, x \rangle + b$ can be derived by

$$w = \frac{1}{2}(X_+ - X_-) = \frac{1}{2}Uv + \mu_+ - \mu_- \tag{8}$$

$$b = -\frac{1}{2}w^T(X_+ + X_-) \tag{9}$$

$X_+ + X_-$ can be written as

$$K = (U_+ \ U_-) \quad H = \begin{pmatrix} v_+ \\ v_- \end{pmatrix} \tag{10}$$

$$b = -w^T(KH + \mu_+ - \mu_-) \tag{11}$$

LMC classifies through this decision function and puts each feature of positive and negative examples into the decision function for calculation until all the features are put into it, and the maximum accuracy is found.

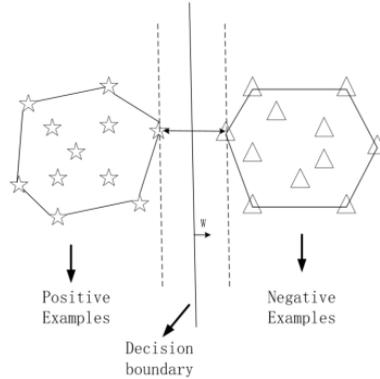


Fig. 1. Two closest points of affine hulls determine decision boundary.

4 Experiment

4.1 Optimized large margin classifier

The original LMC is the feature vector after the PCA dimensionality reduction from the first to the last. Although the energy of the feature vector after PCA processing is the largest at the beginning, the largest amount of energy does not mean the strongest discrimination. In order to find the feature vector with the highest classification accuracy, the original LMC needs to find the last one from the first feature vector, which is very time consuming, and the final accuracy rate is not the highest. Our method is to take one eigenvector from the positive examples and the other from the negative examples each time, and calculate the two eigenvectors with the highest discrimination rate and put them in the next operation in turn.

4.2 Datasets introduction

In this experiment, we have used three datasets, namely MNIST, Fashion-MNIST, Med-MNIST. Among them, MNIST is the easiest to classify and has the highest accuracy rate, and Med-MNIST is the most difficult to classify and has the lowest accuracy rate. We selecte the two most difficult ones to classify for each data set for experiment. The figure below shows three data sets.



(a) MNIST-0

(b) MNIST-6

(c)MNIST-0

(d) MNIST-6

Fig. 2. MNIST dataset.



(a) T-shirt (b) Pullover (c) Sneaker (d) Ankle boot

Fig. 3. Fashion-MNIST dataset.

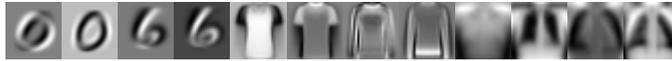


(a) Chestmnist (b) pneumoniast (c) organmnist_axial (d) organmnist_coronal

Fig. 4. Med-MNIST dataset.

4.3 Visualization after PCA dimensionality reduction

We use PCA to reduce the dimensionality of the original images by means of feature normalization to display the feature picture as shown in Figure 5.



(a)Mnist-0 and Mnist-6 (b)Fashion-MNIST-Pullover and shirt (c)Med-MNIST-chest and pneumonia

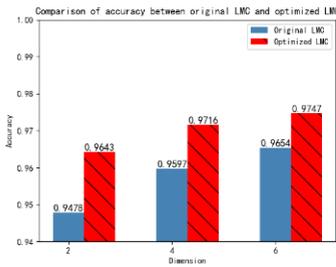


(d)Mnist-8 and Mnist-9 (e)Fashion-MNIST-Ankle boot and Sneaker (f) Med-MNIST-axial and coronal

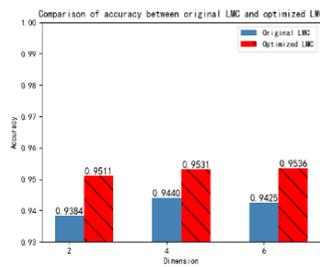
Fig. 5. Visualization after PCA dimensionality reduction.

4.4 Comparison of LMC and optimized LMC

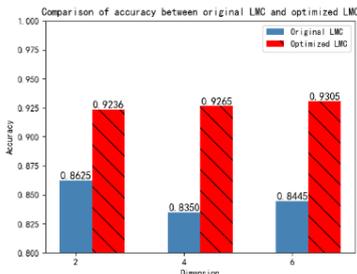
We compare LMC and the optimized LMC in the same dimension. The histogram below is their comparison of accuracy.



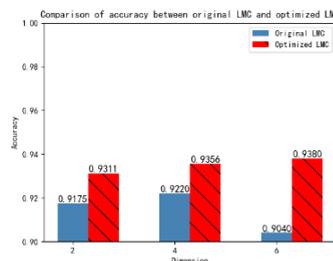
(a) MNIST-0 and Mnist-6



(b) MNIST-8 and MNIST-9



(c) Fashion-MNIST-Pullover and shirt



(d) Fashion-MNIST-Ankle boot and Sneaker

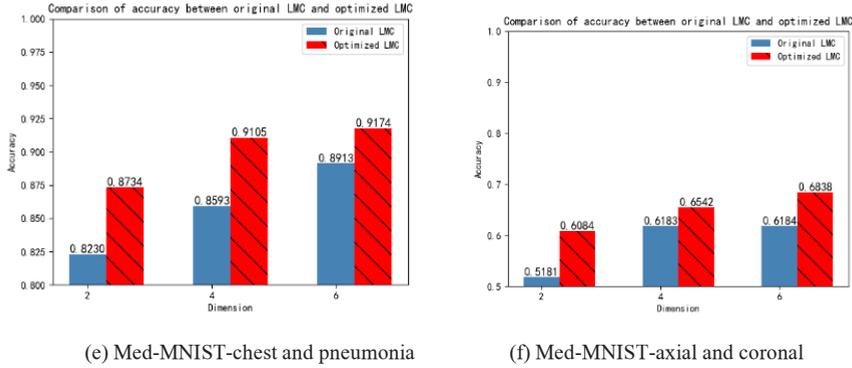


Fig. 6. Comparison of accuracy between original LMC and Optimized LMC.

5 Conclusion

From Figure 6 we can see that in the same dimension, the accuracy of the optimized LMC is higher than that of the original LMC. We have found that the accuracy of our optimized LMC starts to increase slowly in most cases after taking three columns of feature vectors for positive and negative examples. In some cases, when the optimized LMC takes two eigenvectors, its accuracy is higher than when the original LMC takes six eigenvectors. In general, our optimized LMC is more accurate than the original LMC algorithm, and our proposed optimization algorithm is feasible.

Supported by Natural Science Foundation of China(Grant No. 6207024186)

References

1. Hakan Cevikalp, Bill Triggs, Hasan Serhan Yavuz, et al. Large margin classifiers based on affine hulls. 2010, 73(16):3160-3168.
2. S. Rajendran, A. Kaul, R. Nath, A. S. Arora and S. Chauhan, "Comparison of PCA and 2D-PCA on Indian Faces," 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014), Ajmer, 2014, pp. 561-566, doi: 10.1109/ICSPCT.2014.6884932.
3. X. Liu and J. Tang, "Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method," in IEEE Systems Journal, vol. 8, no. 3, pp. 910-920, Sept. 2014, doi: 10.1109/JSYST.2013.2286539.
4. Y. Yang, J. Wang and Y. Yang, "Improving SVM classifier with prior knowledge in microcalcification detection1," 2012 19th IEEE International Conference on Image Processing, Orlando, FL, 2012, pp. 2837-2840, doi: 10.1109/ICIP.2012.6467490.
5. J. Mrva, Š. Neupauer, L. Hudec, J. Ševcech and P. Kapec, "Decision Support in Medical Data Using 3D Decision Tree Visualisation," 2019 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 2019, pp. 1-4, doi: 10.1109/EHB47216.2019.8969926.
6. Y. Wang, Y. Ou, X. Deng, L. Zhao and C. Zhang, "The Ship Collision Accidents Based on Logistic Regression and Big Data," 2019 Chinese Control And Decision

- Conference (CCDC), Nanchang, China, 2019, pp. 4438-4440, doi: 10.1109/CCDC.2019.8832686.
7. X. Liu and J. Tang, "Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method," in *IEEE Systems Journal*, vol. 8, no. 3, pp. 910-920, Sept. 2014, doi: 10.1109/JSYST.2013.2286539.