Classification using Different Normalization Techniques in Support Vector Machine

Priti Sudhir Patki
EXTC Department
Dwarkadas J. Sanghvi College of Engineering
Vile-Parle (West), Mumbai

Vishakha V.Kelkar
EXTC Department
Dwarkadas J. Sanghvi College of Engineering
Vile-Parle (West), Mumbai

ABSTRACT
Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, data classification etc. The Support Vector Machine is a popular classification technique. In this paper we have performed different normalization techniques on different datasets. These techniques help in obtaining high training accuracy for classification. The classification is performed on these datasets using SVM.

Keywords
Classification, Normalization, Support Vector Machine, Kernel Functions.

1. INTRODUCTION
The Support Vector Machine (SVM) first proposed by Vapnik has attracted a high degree of interest in the machine learning research community. Several recent studies have reported that the SVM generally are capable of delivering higher performance in terms of classification accuracy than the other data classification algorithms. Classification is an important and widely used technique in many disciplines, including statistics, artificial intelligence, operations research, computer science and data mining and knowledge discovery. Before using classification algorithms, pre-processing operations is one of the important things that should be done to improve the accuracy of classification algorithms. Pre-processing operations include various methods, one of them is normalization.

In this paper, we have used three different datasets for determining the accuracy for classification i.e. Heart data, Seeds data and Iris data from the UCI repository. All these datasets have different number of training data, testing data, feature vectors and classes. Different normalization techniques are performed on these different datasets. And then the accuracy of classification algorithm is calculated before and after normalization on these datasets. In this study, the SVM algorithm is used in classification since this algorithm works based on n-dimension space and if the data sets become normalized the improvement of results will be expected.

The paper is organized as follows. In Section 2, we have explained in detail about Support Vector Machine (SVM) classification. In Section 3, we have given a brief description of different normalization techniques used. Section 4 describes the analysis conducted by us along with the experiment results. This is followed by the conclusion in section 5.

2. SVM CLASSIFICATION
Support Vector Machine (SVM) [1], [2], [3], [4] is a supervised learning model with associated learning algorithms that analyse data and recognize patterns, used for classification and regression analysis [5] of multispectral [6] satellite images. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making it a non probabilistic binary linear classifier.

An SVM model is a representation of points in space which are mapped into separate categories divided by a clear gap as wide as possible. In addition to performing linear classification, SVMs can efficiently perform non-linear classification using kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

2.1 Linear SVM

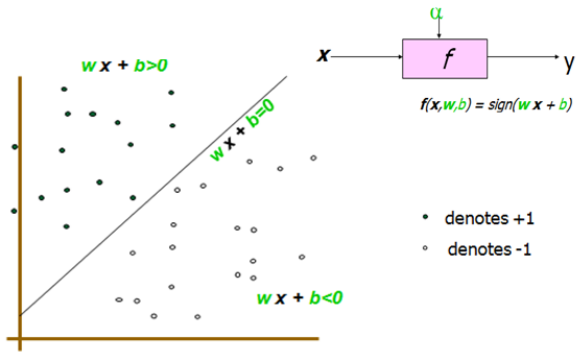

Fig 1: Linear SVM classifier

A linear SVM classifier [7] is as shown in fig 1. A two-class classification problem can be stated as follows: N training sample are available and are represented by the set pairs {(yi, **x**i), i = 1, 2, …, N} with yi a class label of value ±1 and **x**i ∈ R^n feature vector with n components. The classifier is represented by the function f(**x**; **α**) → y with α as the parameter of the classifier.

The SVM method finds the optimum separating hyper plane such that:

1) Samples with labels $y = \pm 1$ are located on each side of the hyper plane;

2) The distance of the closest vectors to the hyper plane on each side should be maximum. The closest vectors are called support vectors and the distance is the optimal margin.

The hyper plane is defined by w.x + b = 0 where (w, b) are the parameters of the hyper plane. The vectors that are not on this hyper plane lead to either w.x + b > 0 or *w.x + b* < 0 and allow the classifier to be defined as: f(x;α) = sign(w.x + b).

The support vectors lie on the two hyper planes, which are parallel to the optimal hyper plane. The equation of the two hyper planes given by $w.x + b = \pm 1$.
The maximization of the margin with the equations of the two support vector hyper planes leads to the following constrained optimization problem:

$$\phi(\omega) = \frac{1}{2}\omega\omega^{T} \text{ is minimized and for all } \{(x_i, y_i)\}, y_i\,(\omega^{T}x_i + b) \geq 1.$$

2.2 Non-linear SVM

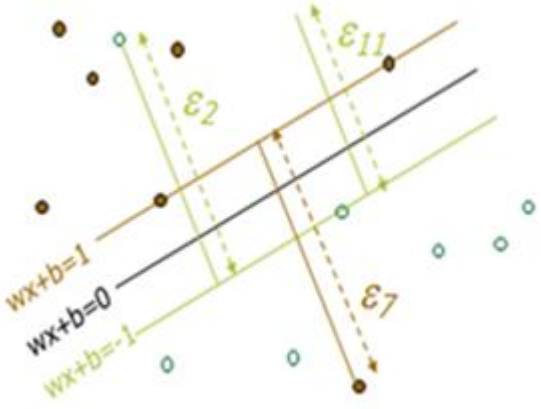

Fig 2: Non-linear SVM classifier

If the training samples are not linearly separable as shown in fig 2, then a non-linear SVM classifier [8] is used in which regularization parameter *C* and error variables εi are introduced in order to reduce the weightening of misclassified vectors. The maximization of the margin with the equations of the two support vector hyper planes leads to the following constrained optimization problem:

$$\phi(\omega) = \frac{1}{2}\omega\omega^{T} + C\sum\xi_i \text{ is minimized and for all } \{(x_i, y_i)\}, y_i\,(\omega^{T}x_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for all i.}$$

The general idea of non-linear SVM is that the original input space can always be mapped to some higher-dimensional feature space where the training set is separable as shown in fig 3.

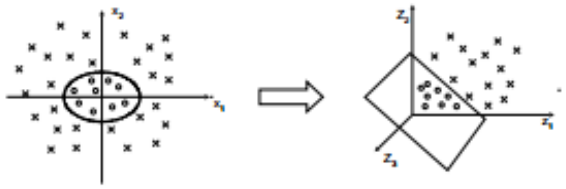

Fig 3: Mapping non-linear data to a higher dimensional feature space

This can be implemented using kernel functions. Kernel functions processes dual maximum margin problem in feature space using linear classification. The resulting model is then a linear model in feature space and a non-linear model in input space. Fig 4 shows the typical SVM algorithm.

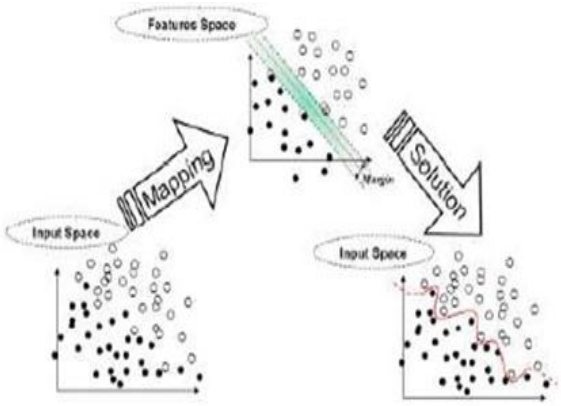

Fig 4: SVM algorithm

However, for general purposes, there are some popular kernel functions [9]:

$$Linear\ Kernel: K(x_i,\ x_j) = x_i^T x_j$$

$$Polynomial\ Kernel: K(x_i, x_j) = (\gamma x_i^T x_j + r)^d\ , \gamma > 0$$

$$RBF\ Kernel: K(x_i, x_j) = exp(-\gamma\ \|x_i - x_j\|)^2\ , \gamma > 0$$

$$Sigmoid\ Kernel: K(x_i, x_j) = \tanh(\gamma\ x_i^T x_j + r)\ , \gamma > 0$$

3. NORMALIZATION TECHNIQUES

Normalization is a method used to standardize the range of independent variables or features of data. It is generally performed during the data pre-processing step. Normalization can be performed at the level of the input features or at the level of the kernel [10].
In many applications, the available features are continuous values, where each feature is measured in a different scale and has a different range of possible values. In such cases, it is often beneficial to scale all features to a common range by standardizing the data.
Different normalization techniques are discussed in this section as follows.

3.1 Linear Normalization (I)

$$r_{ij} = \frac{x_{ij}}{x_j^*}$$

i=1,…., m; j=1,…, n; x_j^*=max$\{x_{ij}\}$

3.2 Linear Normalization (II)

$$r_{ij} = \frac{x_{ij} - x_j^{\sim}}{x_j^* - x_j^{\sim}}$$

i =1,…, m; j=1,…., n; $x_j^{\sim}$ =min$\{x_{ij}\}$

3.3 Linear Normalization (III)

$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}$$

i=1,….,m; j=1,….,n

4. EXPERIMENT RESULTS

In this section we present the experimental results on some datasets from the UCI repository of machine learning databases. From the UCI repository we have chosen three datasets for accuracy analysis i.e. Heart data, Seeds data and Iris data. RBF kernel is used for classification for all these datasets. The experiment was conducted in two parts. In the first part, classification was done on these datasets directly without normalization. Table 1 shows the accuracy observed before the normalization process. During this calculation, RBF kernel was used with the default values of C and γ i.e. 1 and 1/ (number of feature vectors) respectively.

Table 1: Accuracy before normalization process

Sr No.	***DataSet***	***% Accuracy***
1	Heart	50
2	Seeds	95.2381
3	Iris	95

In the second part, different normalization techniques were performed on these datasets with the same values of C and γ as in the first part along with 5-fold cross validation. Table 2 shows the accuracy observed after the normalization process.

Table 2: Accuracy after normalization process

Sr No.	***DataSet***	***% Accuracy***		
		Linear Normalization (I)	***Linear Normalization (II)***	***Linear Normalization (III)***
1	Heart	50	59.2593	74.0741
2	Seeds	95.2381	97.619	97.619
3	Iris	95	96.6667	96.6667

5. CONCLUSION

In this paper, we implemented different normalization techniques on various datasets to improve the accuracy of classification using SVM. Depending upon the different properties of these datasets, these techniques have improved the classification accuracy. From the experimental results, it can be seen that the classification accuracy after normalization is much improved as compared to before normalization for various datasets. Also, an improvement in classification accuracy can be seen for different normalization techniques on these datasets.

6. REFERENCES

[1] C. J. Burges, "A tutorial on support vector machines for pattern recognition," in Data mining and knowledge discovery, U. Fayyad,Ed. Kluwer Academic, 1998, pp. 1–43.

[2] Miloš Kovaevi, Branislav Bajat, Branislav Trivi, Radmila Pavlovi, "Geological Units Classification of Multispectral Images by Using Support Vector Machines", 2009 International Conference on Intelligent Networking and Collaborative Systems, 978-0-7695-3858-7/09.

[3] S.V.S Prasad, Dr. T. Satya Savitri, Dr. I.V. Murali Krishna, "Classification of Multispectral Satellite Images using Clustering with SVM Classifier", International Journal of Computer Applications (0975 – 8887) Volume 35– No.5, December 2011.

[4] Helmi Zuihaidi Mohd Shafri, Affendi Suhaili, Shattri Mansor, "The Performance of Maximum Likelihood, Spectral Angle Mapper, Neural Network and Decision Tree Classifiers in Hyperspectral Image Analysis", Journal of Computer Science 3 (6): 419-423, 2007, ISSN 1549-3636.

[5] Vapnik, V., 1998. Statistical Learning Theory. Wiley Publications, New York.

[6] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," Int. J. Remote sensing, vol. 23, no. 4, pp. 725–749, 2002.

[7] Pabitra Mitra *, B. Uma Shankar, Sankar K. Pal, Pattern Recognition Letters 25 (2004) 1067–1074.

[8] Gr´egoire Mercier and Marc Lennon, "Support Vector Machines for Hyperspectral Image Classification with Spectral-based kernels," IEEE Transactions 2003, 0-7803-7930-6.

[9] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, "A Practical Guide to Support Vector Classification", Dept of Computer Science National Taiwan Uni, Taipei, 106, Taiwan.

[10] Asa Ben-Hur and Jason Weston, "A User's Guide to Support Vector Machines", O. Carugo, F. Eisenhaber (eds.), Data Mining Techniques for the Life Sciences, Methods in Molecular Biology 609, DOI 10.1007/978-1-60327-241-4_13, Humana Press, a part of Springer Science+Business Media, LLC 2010.