International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 1 pp. 306-309 Jan 2018-June 2018 www.csjournals.com

Breast Cancer Detection through Support Vector Machine Classification

[1]Dr Anupam Bhatia, [2]Vikash Khobra
[1]Assistant professor, [2]M.Phil Scholar
CRS, University Jind, Haryana, INDIA

Abstract: Breast cancer is deadly disease that affects all people irrespective of whether they live in developed countries or developing countries. Before the disease all are equal whether rich or poor and it do not have discriminative nature. It is omnipotent and it is pervasive in nature. The detection of breast cancer through machine learning may lead to great help to the medical expert. There are numbers of techniques used for classified the breast cancer data. The data was classified by using the Linear kernel, Polynomial kernel, Redial basis function and Sigmoid kernel. It was also found that the accuracy with linear function is highest that is 84.511%. Therefore, detection of the breast cancer SVM should be used for classification with linear function.

Keywords: Support vector machine, kernel function, and Relevance Vector Machine.

I. Introduction

Our healthcare sector daily collects the large amount of data which including investigation reports, clinical examination, vital parameters, treatment follow up and drug decisions etc[1]. But it is not analyzed properly, mine and classification in an appropriate way. It is essential to interpret the correct classification of breast cancer data. Different techniques for classification the breast cancer data like decision, SVM, RVM, Bayesian classification, and neural networks[2]. But SVM is a better data mining techniques for properly classifying the breast cancer data[3].

The major problem is how complex patterns can be presented and bogus patterns are excluded. For this purpose, classification is an efficient solution[4]. Support Vector Machine is a new and learning method used for classification, regression. Support Vector Machine is a new and learning method used for classification[5].

SVM classifiers work for both of linear and non-linear classes of data by using kernel tricks[6].There are number of the hyper planes that classify the data, a best hyper plane is one that represents the largest separation, maximum-margin and also distance from to the nearest data points on each side is maximized[7]. The higher level of accuracy and also reasonably good speed factor is attracting the analysts towards SVM classification. There are different types of classification methods but the Support Vector machine has good track record in classification [8].

II.Materials and Methods

There are selection of kernels for particular data set is very complicated and also a best choice for the Data Miner Analyst as because Support Vector Machine is a kernel-sensitive in nature[9].Choosing the most appropriate kernel highly depends on the problem at hand because it depends on what we are trying to model and fine tuning its parameters can easily become a tedious and burdensome task. A kernel method is the class of algorithms for the Pattern Analysis and also Classification whose optimum well-known element isSupport Vector Machine (SVM)[10]. The mainly characteristic of Kernel Methods isthat their distinct approach for the problem. The Kernel methods also map data into the higher dimensional spaces in hope that higher-dimensional space the data could become more easily separated as well as better arranged[11].

There are also no restraints on the form of this mapping, which could even lead to infinite-dimensional spaces. This mapping function is also hardly needs to be computed because of a tool called the kernel trick[12].

The kernel trick is mathematical tool that is applied to any algorithm which exclusively depends on dot product between the two vectors[13]. Wherever the dot product is used, then it is replaced by a kernel function. When properly applied, those candidate linear algorithms are transformed into the non-linear algorithms[14].

Those non-linear algorithms are equivalent to their linear originals operating in the range space of a feature space. Here, explain the kernel specification. Slightly Change of value affect heavily on accuracy. There are four types of kernels explained below.

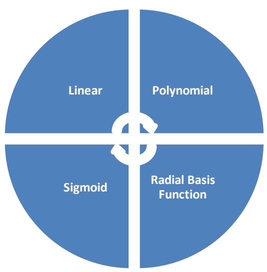

Figure 1.1 Types of Kernels

Linear Kernel

The Linear Kernel is the simplest kernel function. It is given by the inner product <x, y> plus optional constant 'c'. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterparts.

$$k(x, y) = x^T y + c$$

Here, In Figure 1, we are using First Type .i.e. Linear Kernel Function which can be coded as –t 0. If we use it for SVM Multiclass Classification then Accuracy comes out to be 84.511%.

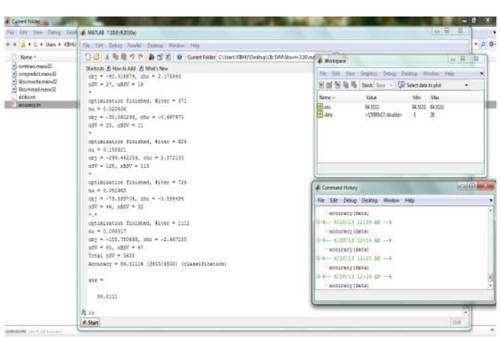

Figure 4.1: Accuracy by using Linear Kernel Function

Polynomial

The Polynomial kernel is a non-stationary kernel. Polynomial kernels are well suited for problems where all the training data is normalized. For degree-d polynomials, the polynomial kernel is defined as

$$k(x, y) = \left(\alpha x^T y + c\right)^d$$

Where xand y are vectors in the input space, i.e. the vectors of features which are computed from the training or the test samples, $c \geq 0$ is constant trading off influence of the higher-order versus the lower-order terms in the polynomial and when $c = 0$, kernel is called homogeneous.

Now,in figure 2, if we will check its Second Type i.e. Polynomial Kernel Function which can be coded as –t 1 using for SVM Classification then its accuracy comes out to be 52.222%.

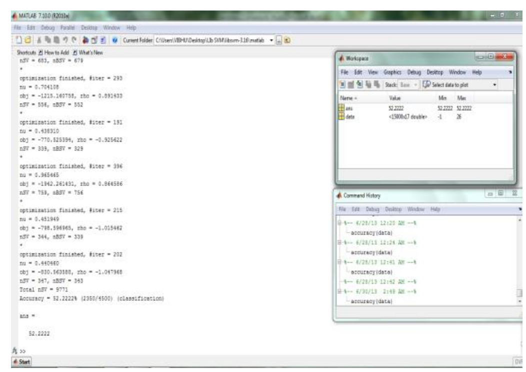

Figure 4.2: Accuracy by using Polynomial Kernel Function

Radial Basis Function

The RBF is the most popular choice of kernel types used in Support Vector Machines. This is mainly because of their localized and the finite responses across the entire range of real x-axis.

There are different types of Radial Basis Kernel Function such as Linear Radial Basis Function, Gaussian Radial Basis Function and Multiquadrics Radial Basis Function.

In Figure 3, for Third Type .i.e. Radial Basis Function which can be coded as –t 2 using for SVM Classification then its Accuracy comes out to be 84.4000%.

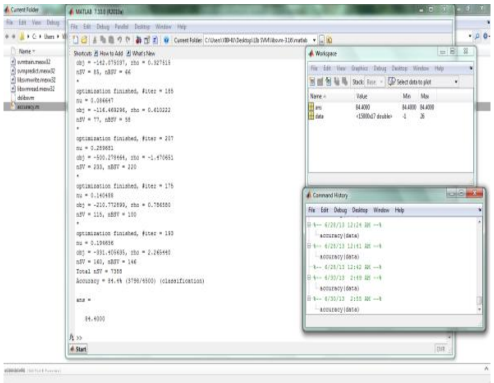

Figure 4.3: Accuracy by using Radial Basis Kernel Function

Sigmoid

The Sigmoid Kernel method is comes from the Neural Networksfield. It is interesting to note that SVM model is using the sigmoid kernel function is equivalent to a two-layer, perceptron neural network. Therefore,kernel method was popular forthe Support Vector Machines due to origin fromneural network theory and despite being only conditionally the positive definite, it has been found to perform well in practice.

$$k(x, y) = \tanh\left(\alpha x^T y + c\right)$$

Sigmoid kernelmethodhas two adjustable parameters in the, the slope alpha and intercepts constant c.

Last but not least, In Figure 4, Fourth Type .i.e. Sigmoid Function which can be coded as –t 3 using for SVM Classification then its Accuracy comes out to be 80.0889%.

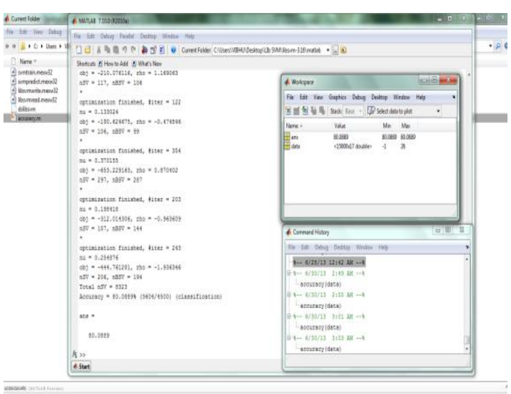

Figure 4.4: Accuracy by using Sigmoid Function

III.Result

In this research paper, there are used different kernel methods i.e. Linear Kernel, Sigmoid, Polynomial, Radial Basis Function which provide classified properly the breast cancer data. But linear kernel provides more accuracy then other kernel methods that Iteration is defined of accuracy rate show as below figure.

optimization finished, #iter = 624

nu = 0.155521

obj = -299.442259, rho = 2.372100

nSV = 128, nBSV = 115

*****

optimization finished, #iter = 724

nu = 0.051963

obj = -79.599706, rho = -0.594494

nSV = 46, nBSV = 32

.

optimization finished, #iter = 1112

nu = 0.093317

obj = -155.780698, rho = -2.467155

nSV = 80, nBSV = 67

Total nSV = 5635

Accuracy = 84.5111% (3803/4500) (classification)

ans = 84.5111

Figure 4.5Iteration Description

Terms used in these Iterations are defined in the following way:

- obj is used for the optimal objective value of dual SVM problem.

- rho is used for bias term in the decision function sgn (w^Tx - rho).
- nSV and nBSV both are the number of support vectors and bounded support vectors (i.e., alpha_i = C).
- nu-svm is used for the somewhat equivalent form of C-SVM where C is replaced by nu and nu simply shows the corresponding parameter.

IV.Conclusion

The accuracy of the breast cancer detection through SVM is calculated using four different linear kernels. The following conclusion is drawn.

- With linear function the accuracy is 84.511%.
-
- With Polynomial kernel the accuracy is 52.222%.
- With Redial Basis Function the accuracy is 84.4000%.
- With Sigmoid the accuracy is 80.0889%.

It is concluded that the breast cancer data is linear in nature. Hence, SVM should be used to detect Breast Cancer Through linear kernel.

Future Work

In future, research can extended to detect other disease through SVM classification.

REFERENCES

[1] Ng, A. (2000). 3 Margins : Intuition. Intelligent Systems and Their Applications IEEE, pt.1(x), 1–25. https://doi.org/10.1016/j.aca.2011.07.027

[2] Liu, H. X., Zhang, R. S., Luan, F., Yao, X. J., Liu, M. C., Hu, Z. D., & Fan, B. T. (2003). Diagnosing breast cancer based on support vector machines. Journal of Chemical Information and Computer Sciences, 43(3),900-907.https://doi.org/10.1021 /ci0256438

[3] Zafiropoulos, E., Maglogiannis, I., & Anagnostopoulos, L. (n.d.). A Support Vector Machine Approach to Breast Cancer Diagnosis and Prognosis.

[4] Wang, Y., & Fuyong, W. (2006). Breast Cancer Diagnosis via Support Vector Machines. Control Conference, 2006. CCC 2006. Chinese,1853–1856.https://doi.org/10.1109/CHICC.2006.280871

[5] Polat, K., & Güneş, S. (2007). Breast cancer diagnosis using least square support vector machine. Digital Signal Processing, 17(4), 694–701. https://doi.org/10.1016/j.dsp.2006.10.008

[6] Shitong, W. (2007). A Novel SVM andIts Application to Breast Cancer Diagnosis.

[7] ??beyli, E. D. (2007). Implementing automated diagnostic systems for breast cancer detection. Expert Systems with Applications, 33(4), 1054–1062.https://doi.org/10.1016/j.eswa.2006.08.005

[8] Qizhong, Z. (n.d.). Gene Selection and Classification Using Non-linear Kernel Support Vector Machines Based on Gene Expression Data.

[9] Purnami, S. W., Rahayu, S. P., & Embong, A. (2008). Feature selection and classification of breast cancer diagnosis based on support vector machines.

[10] Huang, C.-L., Liao, H.-C., & Chen, M.-C. (2008). Prediction model building and feature selection with support vector machines in breast cancer diagnosis. Expert Systems with Applications, 34(1), 578–587. https://doi.org/10.1016/j.eswa.2006.09.041

[11] Liu, J. (2008). A Novel Mixture Classifier and its Application in Breast Cancer Prognosis, 564–568.

[12] Lu, X., Lu, X., Wang, Z. C., Iglehart, J. D., Zhang, X., & Richardson, A. L. (2008). Predicting features of breast cancer with gene expression patterns. Breast Cancer Research and Treatment,108(2),191–201.http://doi. org/10.1007/s10549-007-9596-6

[13] Maglogiannis, I., Zafiropoulos, E., & Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. Applied Intelligence,30(1),24–36.https://doi.or g/10.1007/s10489-007-0073-z

[14] Osareh, A. (2010). Machine Learning Techniques to Diagnose Breast Cancer, 114–120.