

Support Vector Machine - Simple Analysis on Classifying Data, Confusion Matrix and Kernel Selection

Jayasudha.J¹

Assistant Professor, Department of Computer Science,
Kovai Kalaimagal College of Arts and Science, Coimbatore, India¹.

Abstract: Machine learning algorithms come under different categories like supervised, unsupervised and Reinforcement learning. In each case many algorithms are there that serves specific purposes like Regression, Classification, Clustering, and Dimensionality Reduction and so on. Regression type algorithms are mainly used for prediction, Classification is used for dividing or categorizing, Clustering is for grouping similar items. In Reinforcement learning it analyses and optimizes the behaviour of an agent based on the environment. We have to choose the correct algorithms available based on the output required. Because for a single problem we may have different solutions but the thing is we have to choose the one with high accuracy and it should fit for any updating in the datasets.

Keywords: Machine Learning, Classification, Kernel, Hyper Planes and Confusion Matrix.

I. CLASSIFICATION ALGORITHMS IN MACHINE LEARNING

This Classification is also one type of Regression that is used to separate the data according to the categorical variable. Classification is used in applications like Image classification and Fraud Detection etc..This classification may be Binary classification or Multiclass Classification. In binary class classification it deals with only two values 0 or 1. In case of multiclass it supports more than 2 variables. Classification model tries to conclude its results from the input.[1] Many algorithms are available in machine learning for classification includes Logistic Regression, K-Nearest Neighbourhood, Naïve Bayes and Support Vector Machine and so on. The entire above classification algorithm serves with different purposes. Because based on the size of the dataset we have to choose the algorithm to classify the dataset into different classes. Here I have taken Support vector machine for discussion as it works and builds model with small datasets.[8] In Machine learning we have different algorithms for classification like logistic regression which is used for binary classification with hypothesis and sigmoid curve, Naïve bayes is used for multi class classification and it usually requires very less training data, KNN algorithm is also one of the most important algorithm with lot of applications like intrusion detection etc. Random forest is used for both regression and classification and this is used for eliminating the over-fitting of a dataset.

II. INTRODUCTION TO SUPPORT VECTOR MACHINE

Support Vector Machine also called as SVM is used for classifying both binary and multi-class variables.SVM is one of the best classification algorithms if the available dataset is too small with nearly 100+ records. As the data is small it shows a better result. SVM's works is to separate the data based on the category for that it is using the line.[1] This line separates the data .Data used in our dataset may be linear or non-linear. Consider we are having two categories of data, if that data is separated without any collision then the data is linearly separable, if not it is non-linearly separable as in the second figure. This separation is done by decision Boundary line called hyper plane, which separates the data clearly with equal distance. In the Fig 2.1 it shows linearly separable data A and B, which is separated through a line and non-linearly separated data which shows the combination of two classes A and B.

III. HYPER PLANES AND DECISION BOUNDARIES

A Decision Boundary that separates the data clearly is called Hyper plane, which separates the different classes of data in equal distance.[2] More Hyper planes can be drawn, but the best hyper plane is the one which maintains equal distance from two classes as shown in the Fig 3.1.

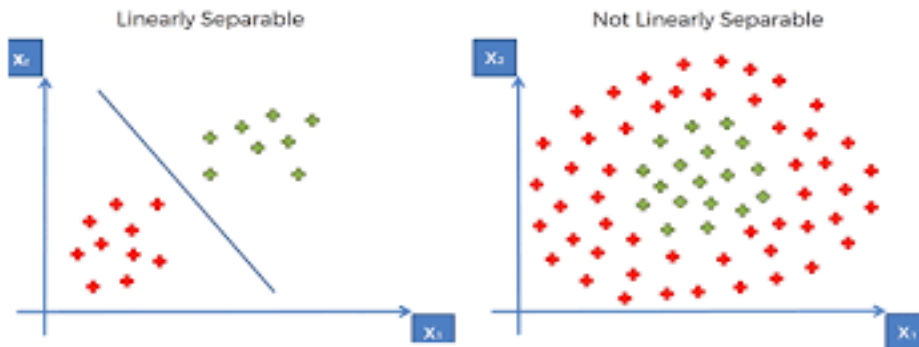


Fig.2.1 Linear and Non-Linear Data

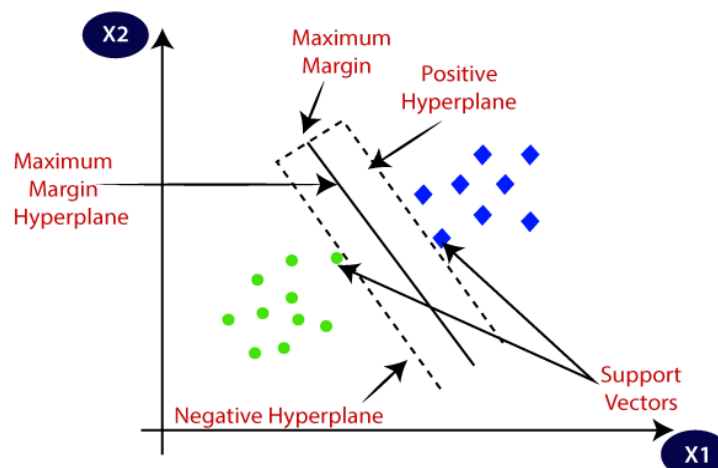


Fig 3.1 View of a Hyper Plane and Margins

It shows a hyper plane with two margins which is available in dotted lines, which still makes our classification more accurate. To support the hyper plane we are adding positive hyper plane and negative hyper plane that maximizes its boundary as large as possible. . The distance between positive and negative hyper plane is called Maximum margin. The points lying on that dotted line are called support vectors.

IV. SVM FOR DIFFERENT DIMENSIONS

While building a model we may consider several aspects. So it is unable to process the figure with least inputs like Single Dimension and Double Dimension to represent the vectors. This will be the case for most of the problems. But for problems with more inputs then we have to move the model for three Dimensions. It is represented in the below Fig 4.1. Usually for non-linearly separable data we have to use this methodology. Here in Lower dimension the data is clumsy and we cannot separate clearly using the hyper plane. Only if it is converted into higher dimension it is separable clearly as shown in the Fig 4.1. So before processing the data, we have to clearly separate the data for further processing.

V. BUILDING MODEL AND CONFUSION MATRIX IN SVM

Choosing the dataset is more important in case of SVM, because it fits only with small datasets. So I have chosen a Heart disease dataset from Kaggle which contains nearly 130 records. Here I have compared several kernels like linear, sigmoid, rbf, poly for implementation to know which one is giving accurate result. [6] First the dataset is divided for training and testing. Here I have used 0.20 as test size. Using Sklearn’s SVC classifications are done. After that to know the performance of the dataset, I have used confusion matrix followed by Classification report with Precision, Recall and F1-score. Finally with accuracy score. A performance measurement technique used in classification. It comes up with 4 different values like True Negative, False Negative, True Positive and False Positive.

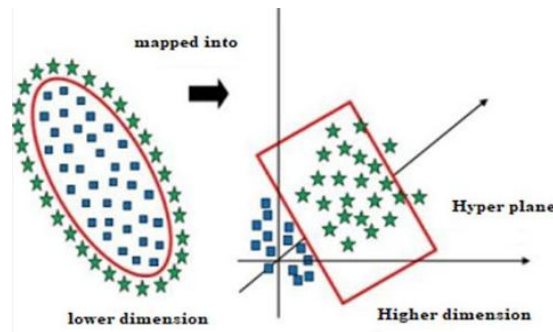


Fig 4.1 Dimensions conversion

This Confusion matrix is for Binary Classification and it increases its size, if we are dealing with multi class classification. That is, it may be 3 x 3 matrix or more. For different types of kernel we used to compare this confusion matrix to know the performance of each kernel.

		Predicted class		Total :
		P	N	
Actual Class	P	True Positives (TP)	False Negatives (FN)	27
	N	False Positives (FP)	True Negatives (TN)	
		20	41	

Fig 5.1 Confusion Matrix & Results from the datasets.

- TP: True Positive: Predicted values correctly predicted as actual positive
- FP: Predicted values incorrectly predicted an actual positive. i.e., Negative values predicted as positive
- FN: False Negative: Positive values predicted as negative
- TN: True Negative: Predicted values correctly predicted as an actual negative.

The above figure shows the result of a confusion matrix with linear kernel for our dataset. This shows the actual values and predicted values. By using this comparison we will be able to choose an appropriate kernel.

VI. WHY CONFUSION MATRIX IS NEEDED?

The above fig 5.1 shows the Resultant confusion matrix that is built using linear kernel implementation. Confusion matrix is built using the Test size (303 * 0.20) = 61 records are used for building confusion matrix. Here TN and TP are considered for evaluation. Because it returns the correct value, true as true and false as false. In the above confusion matrix 19+33 = 52 values are predicted correctly out of 61. [5] So this linear kernel seems to be a better case and it should be compared with other kernels to know the best kernel. With this we are able to calculate the accuracy, precision, True positive rate, True Negative Rate, False Positive Rate, False Negative Rate, Precision and Recall.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$accuracy = 52/61 = 0.85$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = 33 / 34 = 0.97$$

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = 33 / 41 = 0.80$$

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

$$F_1 = 2 * ((0.80 * 0.97) / (0.80 + 0.97))$$

$$F_1 = 0.88 \text{ (Rounded off)}$$

In this way confusion matrix calculates the values for various measures for different kernels like sigmoid, poly, linear, rbf and so on.

A. Choosing a correct kernel for our classification problem.

In SVM it uses different mathematical functions that are named as kernel.[7] These functions are of different types like linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. So transforming a lower dimension to a higher dimension is called as a Kernel. [9]Based on the data which may be linear or non-linear,[3] if it is a linear one then we can choose the linear kernel function which can separate data using a single line, if the data is non-linear then we have to choose some other kernel methodologies[4] like rbf, poly and so on.

B. How do we know the data is linear or non-linear?

This can be found using graphs or plots. With the help of various plots/graphs, we can make a graph. Based on the structure of the graph we can choose the linear one or non-linear one. And if we are dealing with binary class variables we are able to plot data in two colors and if we are dealing with multiclass variable we may choose various colors for different values. For building graph like structures we are having different libraries in python which shows a clear view of data scattering. Because if we are having non-linear data then, it should be converted into multi dimensional model to make them organized. Then only we will be able to clearly separate the data like linear one.

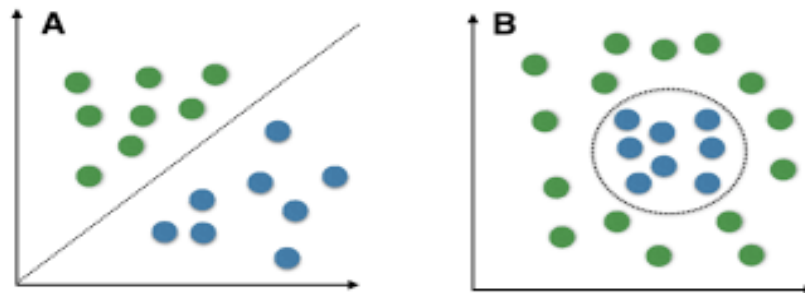


Fig 6.1 Linearly Separable and non-Separable.

In the above Fig 6.1 figure A, data is linearly or clearly separable, but in figure B data cannot be separated by a single line like A. In this case we cannot use the linear kernel, we have to choose the some other kernel which suits the model.

VII. SVM APPLICATIONS IN REAL WORLD

SVM usually works better in higher dimensions and with normal datasets. And this is best suited for binary classification problems. If we are dealing with larger dataset then its performance doesn't seem good. We feel difficult to select an appropriate kernel for our problem. SVM is a supervised machine learning algorithm and it is mainly used for classification or categorizing.

Some real-life applications are:

- Handwriting recognition
- Bioinformatics
- Protein fold and remote homology detection
- Image classification
- Face Detection

VIII. CONCLUSION

As I have taken a small dataset, I used SVM for my model with nearly 130 records. If the dataset is large then this classification model would not give us a best result. So we have to choose some other classification algorithms like Naïve-Bayes, KNN, Decision Tree or it may be Random Forest Algorithm. Our dataset has linear data and it produced various results for different kernels like for Linear it Produced the accuracy level of 0.85, Sigmoid = 0.56 (very poor performance), Poly = 0.69 and rbf = 0.64, here I have compared only these four and among these linear kernel produced the better result and performance. In Machine Learning several algorithms are there, used for different cases. So before selecting an algorithm we should be clear about the problem whether it is regression, classification or clustering type. Because each type many algorithms are available which supports larger or smaller datasets, Binary or multiclass datasets and so on.

REFERENCES

- [1]. Durgesh, K. S., & Lekha, B. (2010). Data classification using support vector machine. *Journal of theoretical and Applied information technology*, 12(1), 1-7.
- [2]. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on Support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215.
- [3]. Apostolidis-Afentoulis, V., & Lioufi, K. I. (2015). SVM classification with linear and rbf kernels. July): 0-7 [http://www.academia.edu/13811676/SVM_Classification_with_Linear_and_RBF_kernels.\[21\]](http://www.academia.edu/13811676/SVM_Classification_with_Linear_and_RBF_kernels.[21]).
- [4]. Patle, A., & Chouhan, D. S. (2013, January). SVM kernel functions for classification. In *2013 International Conference on Advances in Technology and Engineering (ICATE)* (pp. 1-9). IEEE.
- [5]. Santra, A. K., & Christy, C. J. (2012). Genetic algorithm and confusion matrix for document clustering. *International Journal of Computer Science Issues (IJCSI)*, 9(1), 322.
- [6]. Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based Learning methods*. Cambridge university press.
- [7]. Hong, H., Pradhan, B., Bui, D. T., Xu, C., Youssef, A. M., & Chen, W. (2017). Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China). *Geomatics, Natural Hazards and Risk*, 8(2), 544-569.
- [8]. Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.
- [9]. Lucieer, A. (2007, July). Visualization of Hyperplanes for SVM classification. In *2007 IEEE International Geoscience and Remote Sensing Symposium* (pp. 2034-2035). IEEE.