Support Vector Machine Technique for EEG Signals

2 authors:

P Bhuvaneswari

Yazhli Global Multidisciplinary Research Organization (YGMRO)

13 PUBLICATIONS   **111** CITATIONS

J. SatheeshKumar

Bharathiar University

61 PUBLICATIONS   **416** CITATIONS

Some of the authors of this publication are also working on these related projects:

Medical Imaging View project

Support Vector Machine Technique for EEG Signals

P Bhuvaneswari
Research Scholar
Bharathiar University
Coimbatore

J Satheesh Kumar
Assistant Professor
Bharathiar University
Coimbatore

ABSTRACT

Support Vector Machine (SVM) is one of the popular Machine Learning techniques for classifying the Electroencephalography (EEG) signals based on the neuronal activity of the brain. EEG signals are represented into high dimensional feature space for analyzing the brain activity. Kernel functions are helpful for efficient implementation of non linear mapping. This paper gives an overview of classification techniques available in Support Vector Machine. This paper also focus role of SVM on EEG signal analysis.

General Terms

EEG Signal Processing.

Keywords

Support Vector Machine, Electroencephalography, classifier, Signal processing.

1. INTRODUCTION

Electroencephalography (EEG) is the prominent technology for identifying the brain abnormalities in many challenging applications in the field of medicine which includes Seizures, Alzheimer Disease, Coma, Brain death, Dysarthria [1]. Paralyzed peoples are not having muscle control, in order to capture the brain signals for analyze the activity of the brain [2].

EEG signals will be generally represented in high dimensional features space and it is very difficult to interpret. Machine learning methods are helpful for interpreting high dimensional feature sets and analyze the characteristics of brain patterns. Support Vector Machine is one of the popular Machine Learning methods for classifying EEG signals. SVM aims to maximize the margin in order to avoid the risk of over fitting data and minimize the misclassification error.

In conventional methods like multilayer perceptron, complexities are controlled depends on number of features used where as in SVM complexities are independent from dimensionality. Optimization problem occurs due to conversion of data into high dimensional feature space and it can be resolved by using inner product of Kernel methods [3].

Organization of this paper includes, a brief discussion about SVM is explained in Section 2. Section 3 deals with SVM classification for EEG signals and section 4 conclude the paper.

2. SUPPORT VECTOR MACHINE

Support Vector Machine was initiated by Vapnik and Cortes for two group classification problem. SVM is applied in many applications like EEG signal classification, cancer identification, bioinformatics, seizure prediction, face recognition and speech disorder.

Support Vector Machine techniques [4] can be classified into three types namely, Linearly separable, Linearly Inseparable and Non linearly separable as shown in Figure 1.
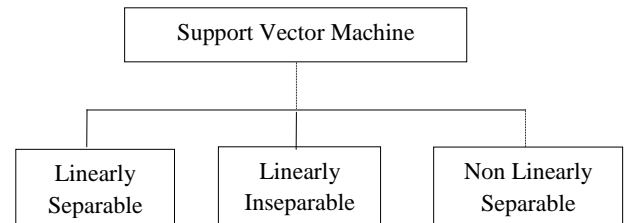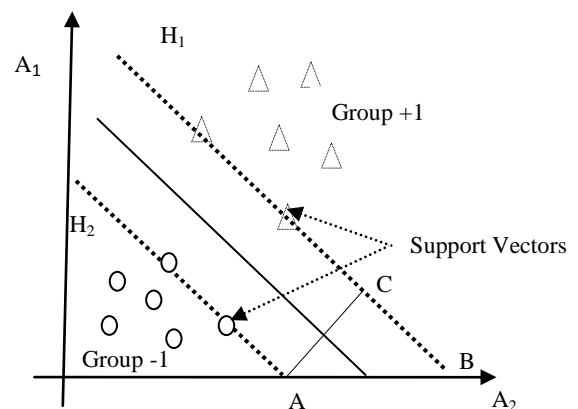

Fig 1: Types of Classification

SVM is used to construct the optimal hyperplane with largest margin for separating data between two groups. For two dimensional data, single hyperplane is enough to separate the data into two groups such as +1 or -1. Two hyperplanes are needed to separate the data points for three dimensional data.

Fig 2: Linearly Separable

SVM constructs hyperplane for separating the sample data based on the target categories. For two dimensional data, there are number of possible linear separators (hyperplanes) and need to find the optimal hyperplane which has maximum margin width. The lines H1 and H2 are drawn parallel to

optimal hyperplane (solid line) and mark the distance between the hyperplane and the data points. The distance between the dotted lines (AC) is called as margin. Some of the sample data points which lie on the hyperplane H1 and H2 are called as Support Vectors (SVs) refer Figure 2. These SVs are essential for calculating the margin width.

2.1 Linearly Separable

Linearly Separable classification seperates the high dimensional data into two groups, $y_i=\{+1,-1\}$ without any overlapping or misclassification (refer Figure 2). SVM produces number of decision margins where the best margin is identified by using perceptron algorithm.

The main objective of Support Vector Machine is maximizing the margin width in order to reduce the misclassification error. The margin width can be calculated by drawing a line (AC) between H1 and H2 and form a triangle ABC.

The optimal hyperplane is given by the equation,

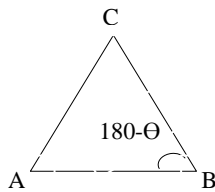$$w_1 x_1 + w_2 x_2 - b = 0 \quad (1)$$

For hyperplane H1 and H2 are represented as,

$$w_1 x_1 + w_2 x_2 - b = 1 \quad (2)$$

$$w_1 x_1 + w_2 x_2 - b = -1 \quad (3)$$

where w_1, w_2 are positions of the hyperplane H1 and H2 respectively. x_1, x_2 are data points and b takes value of +1,0,-1 which shows how far hyperplanes away from the original line.

The distance between two hyperplanes is measured by calculating the length of AC.

C

180-Θ

A   B

AC= - AB sin Θ   (4)

Assume x_2=0, substitute x_2 in (2), get

$$w_1 x_1 = 1 + b => x_1 = \frac{1+b}{w_1}$$

Assume x_2=0, substitute x_2 in (3), get

$$w_1 x_1 = b - 1 => x_1 = \frac{b-1}{w_1}$$

A $\overline{\hspace{3cm}}$ B

$(\frac{b-1}{w_1}, 0)$ $(\frac{1+b}{w_1}, 0)$

By applying Euclidean distance,

$$AB^2 = \left(\frac{b-1}{w_1} - \frac{1+b}{w_1}\right)^2$$

$$AB = -\frac{2}{w1}$$

Substitute AB value in (4)

$$AC = \frac{2}{w_1} sin\theta$$

Substitute the value of $sin\theta$, $sin\theta = \frac{w_1}{\sqrt{(w_1^2 + w_2^2)}}$

$$AC = \frac{2}{w_1} \cdot \frac{w_1}{\sqrt{(w_1^2 + w_2^2)}}$$

$$AC = \frac{2}{\|w\|}$$

The maximum margin width is $\frac{2}{\|w\|}$ and minimum margin width is $\frac{1}{2}\|w\|$ with subject to the constraint

$$y_i(w_1 x_{i1} + w_2 x_{i2} - b) \geq 1 \, for \, i = 1,2 \dots m \, (5)$$

The parameters such as weight vector (w), bias (b), number of support vectors (m) are essential for classification. Margin width can be calcutated from the values of 'w' and 'b' by using optimization methods is called primal problem. The values of the parameter are having larger values then it is diffficult to calculate by primal method. This problem is considered as dual problem where as Lagrange multiplier is applied for calculating values.

Primal problem ⟶ Lagrange Multipliers ⟶ Dual Problem → Classifier
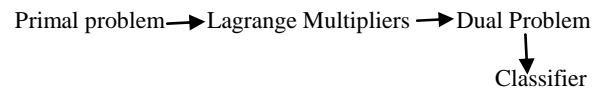
Lagrangian multiplier for primal problem

$$L(w,b,\alpha) = [\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\alpha_i[y_i w'x_i + b] - 1] \quad (6)$$

where α_i is the Lagrangian multiplier

By applying the derivative of L with respect to w and b to zero, get

$$\nabla_w L(w,b,\alpha) = w - \sum_{i=1}^{n}\alpha_i y_i x_i = 0$$

This implies that,

$$w = \sum_{i=1}^{n}\alpha_i y_i x_i \quad (7)$$

Apply w in (6) get

$$L(w,b,\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j x_i' x_j$$

Associated Dual form is

$\max L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i\, y_j\, \alpha_i\, \alpha_j\; x_i'x_j$ with subject to $\sum_{i=1}^{n} \alpha_i\, y_i = 0 \; \forall\, i$

The optimum seperating hyperplane (OSH) can be calculated by quadratic programming(QP). Support vectors can gain the α value and helps to obtain bias value 'b' for training samples.

$$b = \frac{1}{N_{sv}} \sum_{sv \in S} [\, y_{sv} - \sum_{i \in s} \alpha_i\, y_i x_i x_{sv} \,]$$

Finally, optimal decision function of a classifier is defined as

$$f(y) = sign(\sum_{i=1}^{sv} \alpha_i\, y_i(x_i, x_{sv}) + b)$$

2.2 Linearly Inseparable

SVM supports for the data contains errors, noisy and faulty data which having some possibilities of error rate. Its impossible to construct a linear hyperplane without error for binary classification data as shown in Figure 3. Linearly inseparable classification can produce solutions for high dimensional datasets with overlapped or misclassified or erronious data. Slack variable ξ is used to represent the error term with slight modification in constraint (5) and allow misclassified points.

$wx_i - b \geq +1 - \xi_i$ for $y_i = +1$ where $\xi_i \geq 0 \; \forall i$ (8)

$wx_i - b \leq -1 + \xi_i$ for $y_i = -1$ where $\xi_i \geq 0 \; \forall i$ (9)

Combine (5) and (6) get,

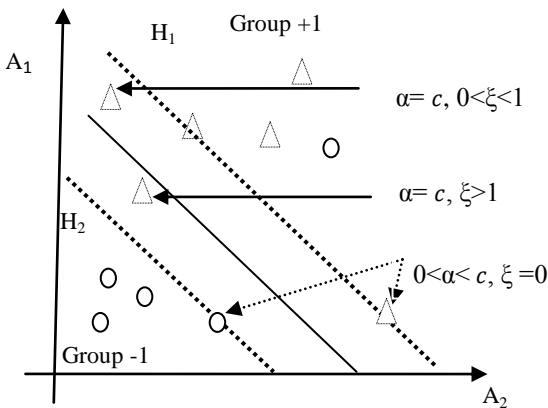$y_i(wx_i + b) \geq \xi_i - 1, \xi_i \geq 0 \; \forall i$ (10)

Fig 3: Linearly Inseparable

Some data may be incorrectly classified and need to adjust the hyperplanes for proper classification. Increasing the margin on either side leads to increase in misclassification error rate. It can be minimized by the function

$$minimize\; \frac{1}{2} \|w\|^2 + c \sum_{i=1}^{n} \xi_i$$

where c is the constant to compromise between the size of the margin and ξ. In Figure.3, Data points will be correctly classified if α_i and ξ_i is equal to zero or α_i is equal to zero and ξ_i value is less than one. Data points will be misclassified, if α_i is equal to c and ξ is greater than one. Data points are consider as Support Vectors, if α_i value must greater than 0, c and ξ is equal to zero.

Lagrangian formuation for dual form

$Q(w, b, \xi, \Lambda, \beta) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \left[(w^T x_i + b) y_i - 1 + \xi_i \right] - \beta i \sum_{i=1}^{n} \xi_i$ (11)

The negative sign is used because objective is to maxmize margin width with respect to α_i, β_i and minimize with respect to, w, b and ξ_i.

Differentiate with respect to w, b, ξ_i and get

$w = \sum \alpha_i y_i x_i$, $\sum \alpha_i y_i = 0$ and $c - \alpha_i - \beta_i = 0$

substitute in (11),

$\max L(\alpha) = \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_i (x_i . x_j) + \sum_{i=1}^{n} a_i$ (12)

with constraint $\sum \alpha_i y_i = 0, 0 \leq \alpha_i \leq c$

The value of $'c'$ can be find using $\frac{1}{c} = \min \{\, 1, 0.05 (\frac{m}{200})^{\frac{3}{2}} \}$

where 'm' is the number of data points and $'c'$ is the constant to compromise between size of the margin and ξ. Decision function for dual classifier can be given as,

$$f(y) = sign(\sum_{sv} \alpha_i\, y_i(x_i, x_{sv}) + b)$$

2.3 Non Linearly Separable

Data are not linearly separable can be converted into higher dimensional mapping for classification (Figure 4). The non linear mapping of original sample data is transformed into higher dimensional mapping is called Feature Mapping and its maping function is denoted as Φ(x_i). Kernel functions are used to find the value of mapping function Φ.

$$x_i^T x_{j=} K(x_i, x_j) = \emptyset(x_i)^T \emptyset(x_j)$$

$K(x_i, x_{j,})$ is called the Kernel functions which is based on the inner product of two varients x_i and x_j . In original space dot product of $x_i . x_j$ is used for calculation and it is converted into higher space can be replaced dot products as kernel function.
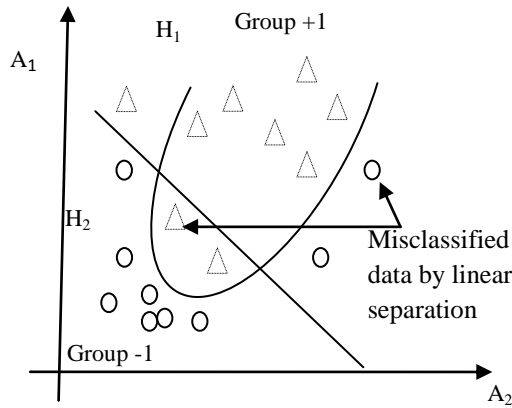
Fig 4: Non Linearly Separable

Some of the popular Kernel functions are as follows;

a) Radial Kernel Function (RBF)

$$K(x_{i,}x_j) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

b) Linear Kernel Function

$$K(x_{i,}x_j) = x_{i,}^T x_j$$

c) Polynomial Kernel Function

$$K(x_{i,}x_j) = \left[(x_{i,}^T x_j) + 1\right]^d$$

d) Gaussian Function

$$K(x_{i,}x_j) = \exp\left(-[\|x_i\text{-}x_j\|^2]/2\sigma^2\right)$$

The optimization function for non linearly separable data is written as,

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{j=1}^{n} a_i a_j y_i y_i \phi_i^T \phi_j$$

Now, move the input data into number of features by replacing with kernel functions

$$\phi_i^T \phi_j \rightarrow K(x_{i,}x_{j,}).$$

$$\max L_d(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_i \ a_i\alpha_j \ K(x_i x_j)$$

Subject to $a_i \geq 0$, i=1…n and $\sum_{i=1}^{n} a_i\ y_i = 0$ where $0 \leq a_i \leq c$ for i=1…n

Finally, decision function is given as,

$$f(y) = sign(\sum_{i=1}^{sv} \alpha_i \ y_i \ k\ (x_i, x_{sv}) + b)$$

3. SUPPORT VECTOR CLSSIFICATION FOR EEG SIGNALS

Support Machine is mainly applied for classifying normal and seizure activity from the continuous recording EEG signals. Feature vectors are generated for both seizure and non seizure activity. RBF kernal function can be choosen as classifier for generating optimal decision boundaries [6].

Feature vectors can be constructed from empirical mode decomposition which decomposes the EEG signal into amplitude and frequency modulated components. Parameters like area and mean frequency of the components are estimated and given as input for LS-SVM. Least square support vector machine with radial basis kernel function can be used for classification of seizure and non seizure activity [7].

EEG signals can be decomposed into different frequency bands with the help of Discrete Wavelet Transform. Features can be generated from entropy, energy and standard deviation and used to train SVM for classifying seizure and non seizure activity [8].

Support Vector Machine classification can be used to classify different kinds of mental tasks like thinking to move left hand, thinking to move right hand, performing mathematical operation and thinking to a carol. Power Spectrum method is applied for extracting features from preprocessed signals and given as training data for SVM. For testing, single channel may be given to classify [9].

SVM also considered for feature selection process. SVM with Gaussian Kernel is used for neonatal feature selection. Various features like mean and standard deviation are extracted from SVM classifier. Training data identifies the optimal hyperplane for neonatal data. [10].

EEG signals can be used to detect the brain tumors with proper testing and training. EEG signal artifacts are removed using adaptive filter method and spectral methods are applied for extracting different spectral bands of frequencies. These features are given to SVM classifier for classifying tumor [11].

SVM can be used to classify emotion detection from EEG signals. Fractal dimension (FD) values of EEG signals are extracted from both hemisphere and allow recognizing emotions with different levels of arousal and valence values. Support vector machine kernel classifier is used to classify the emotions based on positive and negative values of arousal and valence features [12].

4. CONCLUSION & FUTURE WORK

EEG signals are used for identifying various brain disorders and also help the researchers for better understanding of brain. Support Vector Machine is one of the popular machine learning and optimal method for classification of EEG signals. Different kernel function plays a vital role in non linear separable methods.

EEG signals are preprocessed by using different artifact removal algorithms such as ICA, PCA. Features like power spectrum, entropy, mutual information and so on are extracted from preprocessed signals for generating feature vectors. SVM classifier accept feature vector and produce classified signals based on the chosen application. BCI will be focused further.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] S J M Smith, "EEG in the diagnosis, classification, management of patients with epilepsy", J Neurol Neurosurg Psychatry ,Vol.76 (suppl II), pp. ii2-ii7, 2005.

[2] Ki-Hong Kim, Hong Kee Kim, Jong-Sung Kim, Wookho Son, and Soo-Young Lee, "A biosignal based human interface controlling a power wheelchair for people with motor disabilities", ETRI Journal, Vol. 28, No. 1, 2006.

[3] Corinna Cortes, Vladimir Vapnik, "Support Vector Networks", Machine Learning pp.273-297, 1995.

[4] Hyeran Byun and Seong-Whan Lee, "Applications of Support Vector Machine for pattern recognition: A survey"

[5] V.David Sanchez, "Advanced Support Vector Machines and Kernel Methods", Neurocomputing, pp. 5-20, 2003.

[6] Ali Shoeb and John Guttag, "Applications of Machine Learning to Epileptic Seizure Detection", 27th International Conference on Machine Learning, Isreal, 2010.

[7] Varun Bajaj and Ram Bilas Pachori, "EEG Signal Classification Using Empirical Mode Decomposition and Support Vector Machine", Advances in Intelligent and Soft Computing,Vol.131, 2012.

[8] J.R.Panda., " Classification of EEG signals using wavelet transform and support vector machine for epileptic seizure activity", Proceedings of International Conference on Systems in Medicine and Biology, 2010.

[9] Giovanni Costantini, Daniele Casali, Massimiliano and Todisco, "An SVM based classification method for EEG signals", Latest Trends on Circuits, ISSN:1792-4227, ISBN:978-960-474-198-4.

[10] A.Temko, G.Boylan,W.Marnane and G.Lightbody, "Speech Recognition Features for EEG signal description in detection of Neonatal Seisures", 2009.

[11] M.Murugesan and R.Sukanesh, "Towards detection of brain tumor in EEG signals using SVM", International Journal of Computer Theory and Engineering", Vol.1, No.5, 2005.

[12] Olga Sourina and Yisi Liu, "A Fractal based algorithm of emotion recognition from EEG signals using arousal and valence model".