**SUPPORT VECTOR MACHINE**

Shailaja Mysugari

College of Business and Economics

***Abstract: The main aim of this term paper is to describe the Support Vector Machine Algorithm, a supervised model used for classification and regression. The paper focuses on SVM for classification tasks only. There are linear and non-linear classifiers available for Support Vector machine. The paper describes the hard margin and the Soft Margin linear Classifier. The non-linear Classifier is not covered in this paper. The paper also provides the formulations for the data which is linearly separable and not separable. It also describes the properties, applications, advantages and disadvantages of Support Vector Machine.***

Keywords-- Support Vector Machine, Hard Margin, Soft Margin

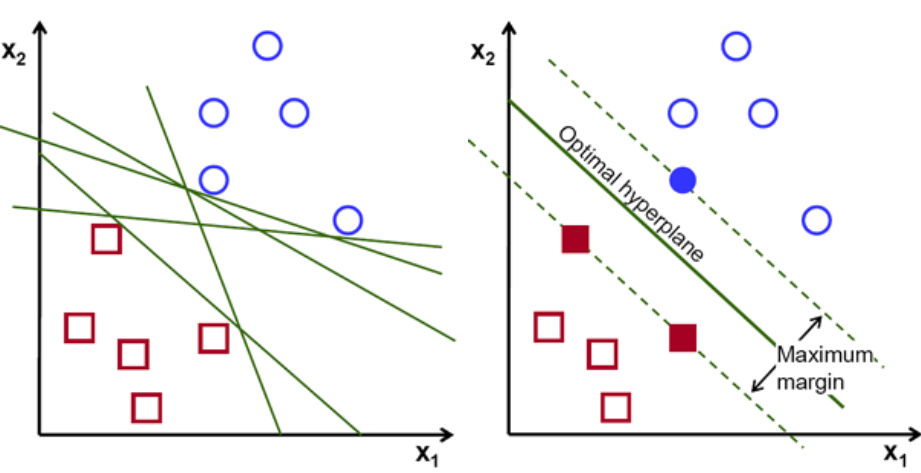
***Introduction***

Support Vector Machine algorithm was originally invented by Vladimir N.Vapnik and Alexey Ya. Chervonenkis in 1963. In the year 1992,Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik  recommended a way to create nonlinear classifiers by applying kernel trick to maximum-margin hyperplanes. The binary class Support Vector Machine was extended to the multiclass (where the number of classes are greater than two )by Weston and Watkins in 1999, Platt in 2000

Support Vector machine(SVM) is one of the most predominant supervised learning algorithms that analyze data and recognize patterns. Support Vector Machines is obtained from Statistical Learning Theory. Support Vector Machine is powerful, yet flexible algorithm used for classification and regression but generally used for classification problems. It is also used for the outlier detection[3]. [Full Article](https://www.researchgate.net/publication/254321762_Recent_advances_on_support_vector_machines_research)

# **WHAT IS THE OBJECTIVE OF SVM?**

The Support Vector Machine objective is to find a hyperplane with N-1 dimensions separating the classes in the N-dimensional feature space. This is called linear classifier. To separate the classes there are many possible hyperplanes. Our goal is to find a hyperplane with N-1 dimension that has maximum margin. The margin is defined as the distance between the separating Hyperplane and the training samples that are closest to this hyperplane, which are so called support vectors. For detailed information on Support Vector Machine see reference [4] [Full Article](https://www.scirp.org/pdf/WSN20100100007_87680380.pdf)

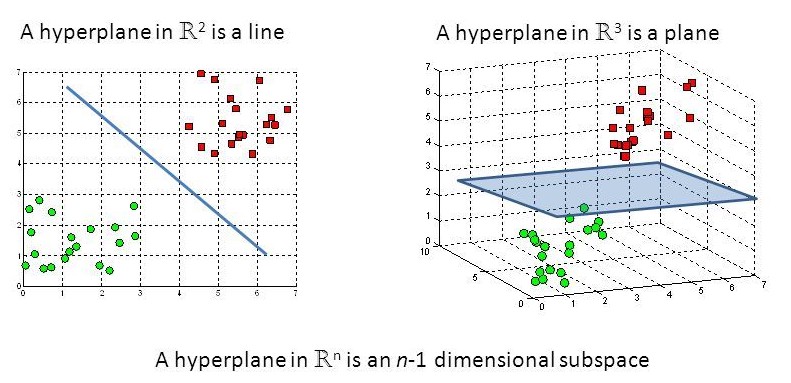
**Infinite Hyperplanes** **Maximum margin Hyperplane** 

**Fig1** Identification of Optimal Hyperplane. [1]

# **HYPERPLANE**

Hyperplane is a subspace which is one dimension less than the original vector space. Let us understand it by examples. For 2 dimensions , a hyperplane is a one-dimensional subspace, in other words it is a line. For 3 dimensions, a hyperplane is a two-dimensional subspace( a plane). So, by generalizing the concept to N- dimensions, a hyperplane is a N-1 dimension subspace. For detailed explanation on Hyperplane, see reference[6] [Source](https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf)

The figure 1 shown below is the hyperplane in two and three-dimensional Space.



**Fig 2** Hyperplanes in 2 D and 3 D space. [1]

In two dimensions, the mathematical formula for hyper plane is given by equation of a line:

w0 + w1x1 + w2x2 = 0

The above concept can be generalized for p dimension

w0+w1x1+w2x2+w3x3+………….+wpxp= 0

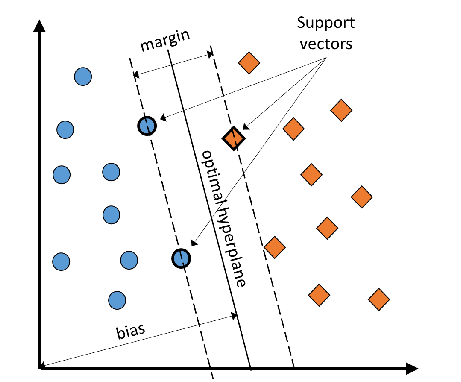
The matrix notation for the above equation

w0+wTx=0

(Application of the property wTx =w.x)

# **SUPPORT VECTORS**

Support Vectors are data points closest to the decision boundary (hyperplane).The separating planes are defined using these so called support vectors. The Hyperplane position and orientation depends on support vectors. If the data point is not a support vector, removing it from the model has no effect. But if the deleted data point is the support vector, it will change the position of the hyperplane. In the figure below the circles represent support vectors of one class and the square represents the support vector from other class. For detailed information, see reference [6] [Source](https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf)



**Fig 3** Support Vectors[2]

# **SUPPORT VECTOR MACHINE -CLASSIFICATION**

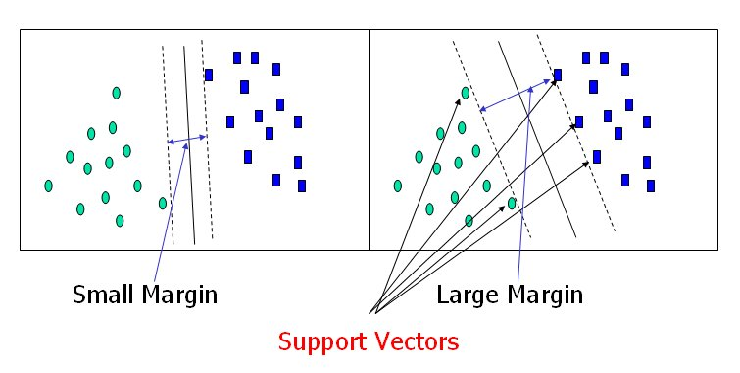
Support Vector Machine is simple and intuitive classifier. There are different types of classifiers available depending upon whether the data is linearly separable or not. If the data is linearly separable by a hyperplane, the maxima marginal classifier also called hard margin Support Vector Machine is used.[8]

## *Hard Margin SVM:*

If the data is linearly separable by a hyperplane then there will be infinite number of separating hyperplanes. Out of all the hyperplanes, the hyperplane with maximal margin( also known as optimal separating Hyperplane) farthest from the training observations is chosen.

What is maximal margin?

The perpendicular distance from each observation in the training dataset to the separating hyperplane is calculated. The smallest such distance is the minimal distance from the observations to the separating hyperplane and is called margin. The maximal margin hyperplane is the one for which the margin is largest. The figure below shows the small margin and large margin with circles representing one class and squares representing another class. Support vectors are the data points used for defining the separating hyperplanes. For detailed explanation on maximal margin ,see reference[6]. [Source](https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf)



**Fig 4**  Hyperplanes with different Margin sizes[1]

Once the optimal separating hyperplane (decision boundary) is chosen, we can then classify the test record depending on which side of the hyperplane it lies.

## *Construction of the Hard Margin SVM:*

For our convenience, let us consider the training data with two classes positive and negative separated by a decision boundary. To get the margin maximization let us look at the positive hyperplanes and negative hyperplanes which can be expressed as follows[7]:

b+wT xpos ≥ 1 -----------------------[1.0]

.

Equation 1.0 defines parallel hyperplanes of positive class

b+wT xneg ≤ ─ 1-----------------------[1.1]

Equation 1.1 defines parallel hyperplanes of negative class

The xpos represents the support vectors for positive class(points on the hyperplane of positive class) and xneg represents the support vectors for negative class(points on the hyperplane of negative class). wT represents the weight vector, b represents the bias or offset

The above equations are based on the concept of signed distance of dot product. The equation [1.0] represents the positive distance and the equation [1.1] represents the signed negative distance from the decision boundary (Hyperplane).

In the above equations we are assuming the samples are classified correctly

Now let us consider label y(i) such that

y(i) is +1 for the positive class

y(i) is -1 for the negative class

The equations [1.0] and [1.1] are multiplied with the label y(i).

Then the final equation

y(i)( b+wTx(i)) ≥ 1………………………[1.2]

y(i)( b+wTx(i)) is called as agreement.

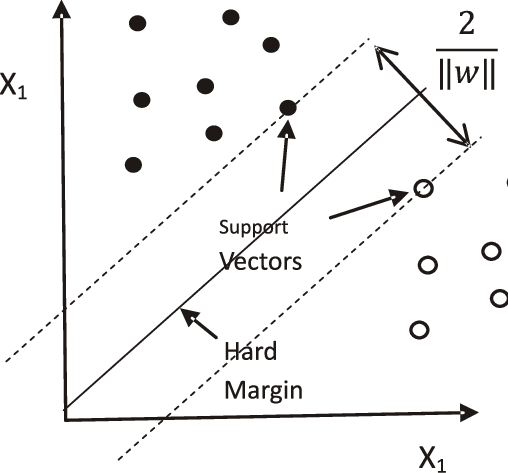
From the above equation we can say that the prediction is correct when y(i) and b+wTx are in same direction.

(For information: The x is representation of support vectors in matrix notation)

Our main aim is to find the maximal margin between two margin boundaries. For that we subtract the equation [1.0] from equation [1.1] to obtain the width (distance between the two parallel hyperplanes)we get

( xpos - xneg) =

The Figure below is a graphical representation of the two parallel hyperplanes



**Fig 5.** Hard Margin SVM. [5]

The left-hand side of the above equation is the distance between the positive and negative hyperplanes which is called margin, the objective to be maximized. Now, the objective is to maximize the . In order to maximize it we need to minimize the 2 which can be solved by using quadratic programming.

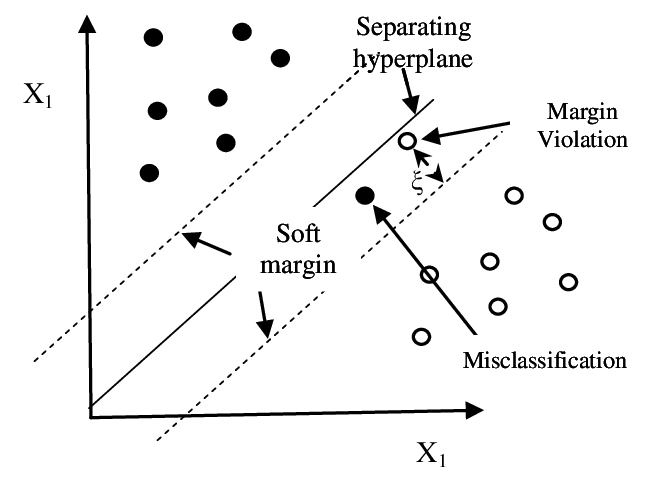
For detailed information about derivation, see reference[7].

The Hard Margin SVM classification is useful only if there is a linear separation of data. But in reality, most of the classification data is not completely separable, so the SVM is extended to Soft margin SVM by introducing a function called Hinge loss function.

## *Soft Margin SVM*

The Soft Margin SVM is the extension of the hard margin SVM for the data which is not separable. SVM, in this case is not looking for maximizing hard margin, instead soft-margin SVM tries to classify most of the data correctly while allowing the misclassification of few points under the appropriate cost penalization [6]. [Source](https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf)

The figure below is a plot of the data with two classes , one class is represented by filled circles and the other by unfilled. It shows the Support Vector Machine in non-separable case



**Fig 6.**  Soft Margin SVM.[5]

The data is not separable in this case, a slack variables ξ(i) is used in SVM objective function to allow error in the classification.(to allow misclassification)

## *Construction of Soft Margin SVM*

The equations for the soft margin SVM are obtained by applying slack variable ξ(i) to the linear constraints. The equations would be same as Hard Margin Support Vector Machine. But we introduce a Hinge loss , allowing contraints to be violated and integrate actual loss value. Let us consider

z= y(i)( b+wTx(i))………………………..[1.2]

The Hinge loss Function is given by

h(z) = 0 if z ≥1

1 – z if z<1…………………..….[1.3]

h(z) is called as Hinge loss.

z greater than or equal to 1 represents the record has been classified correctly and so the loss term is zero. When the z is between 0 and 1 the records are classified correctly with hinge loss less than 1. When the z is less than 1 there will be misclassification, so in this case loss function is given by equation [1.3]

Our goal is to find the maximal hyperplane by allowing some amount of misclassification. So, let us take an example of a dataset with size n and m features. There are two classes in the dataset positive and negative class. The SVM classifier will try to minimize the following function

]

The above function can be divided in to two parts

is a regularization term and is the regularization parameter.

] is the loss term

For convention lambda is removed from regularization term and the loss term is multiplied by variable C ( a misclassification parameter) and term is replaced by h(z) (hinge loss function)

Then our new optimization problem is to minimize the function

+

Subject to constraint

y(i)( b+wTx(i)) ≥ 1-ξ(i) ξ(i)≥0

i=1,2,3……..n

The relation between lambda and C is ƛ =

The variable C is to control the penalty for the misclassification error. Large values of C correspond to the large error penalties whereas for small values of C corresponds to small error penalties. In other words, small value of C means we are less strict about misclassification errors. The value of parameter C is used to control the width of the margin and the bias-variance tradeoff. Increasing the C value decreases the training error. For detailed information ,see reference [7]

Using Gardient Descent and Stochastic Gradient Descent to find optimised solution for the objective function of Support Vector Machine.

The cost Function is given by

J(w,b)= +]

The weight is updated using the Gradient Descent algorithm as follows.

is the learning rate.

For Stochastic Gradient descent a point is

selected randomly from the points i=1,2…..n

The weight is updated using Stochastic Gradient

descent algorithm as follows.

Thus the optimised solution is obtained using the

Gradient descent, Stochastic Gradient descent.

Using Lagrange Multipliers

Introducing the Lagrange Multipliers αi for the constraints to obtain optimal solution , i=1,2,3………………………………N

Minimize Lp(w,b,α) = +

Subject to

This is called the primal problem also referred as Primal Lagrangian. For detailed information see information[8]

***Formulating the dual objective***

Take the partial derivatives of Lp with respect to w set to zero

Take the partial derivatives of Lp with respect to b set to zero

Substituting them in the primal Lagrangian Lp gives the dual Lagrangian

LD(w,b,α) =

-

subject to constraint

0≤ αi≤ C

This is called dual objective function and this function must be maximized for positive values of alpha to find the optimum Hyperplane[8]. .[Full Article](http://www.jatit.org/volumes/research-papers/Vol12No1/1Vol12No1.pdf)

Note: The above two are linear classifiers, there are nonlinear classifiers available in Support Vector Machine using kernel Function. They are popular classifiers ,useful when the data is not linearly separable in input space. For moreodetails on kernel with example ,see reference[2] [Full Article](https://www.researchgate.net/publication/332427600_Development_of_a_Real-Time_Objective_Gas-Liquid_Flow_Regime_Identifier_Using_Kernel_Methods).

## *MULTICLASS SVM*

The Support Vector Machine is extended to multi class classification where the number of classes are more than two. There are different strategies used to address the multiclass classification, they are one versus the rest and pair-wise classification .

One Versus the rest called as one-against -all is one of the most common multiclass SVM used . In this, the algorithm constructs the n binary classifiers where the n is the number of classes in the data. Each classifier discriminates one class from the others, thus reducing it to a binary class(two class) problem. There will be n decision functions in this case. The references are provided below for detailed explanation [23].

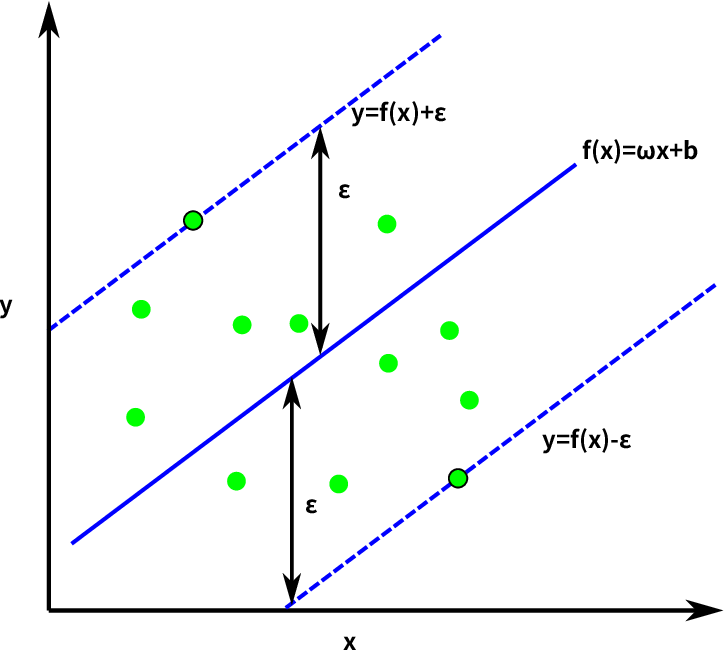
Pair wise classification also called as one-against-one classifier. It constructs n(n-1)/2 binary support vector machine classifiers where n is the number of classes in the dataset. Each classifier distinguishes two of the n classes and it requires evaluation of (n-1) classifiers[23].

The multiclass SVM discussed above sometimes may not fit across the entire dataset. In order to overcome this problem, the data is sometimes divided into subgroups with similar classes and then their features are derived separately. This process results in multistage Support Vector Machine or the Hierarchical SVM which can produce greater accuracy in generalization and also reduces the overfitting problem.[23] [Full Article](https://www.researchgate.net/publication/29687855_Analysis_of_Multiclass_Support_Vector_Machines).

For detailed explanation on the multiclass SVM ,see reference[23] .

# **SVM REGRESSION**

Support Vector Machine is also useful for regression. It supports linear regression as well as non-linear regression. The principle involved in SVM regression is almost similar to the principle in classification with small differences. In regression , consider two lines at a certain distance one above and the other below the hyperplane. The distance between these lines is the margin of tolerance(epsilon) or maximum error. The figure below shows the Support Vector Machine for Regression.[11]



**Fig 7**. SVM Regression for Hard Margin Problem[11]

In the above constraints ,the is the margin of tolerance. The decision boundary are the two dotted lines ,one is above the hyperplane and the other is below it. The data points between these lines are considered.

The objective is to minimize the

2

subject to constraints

y(i)-wx(i)-b ≤

wx(i)+ b -y(i)≤

When the slack variable is considered , the objective becomes to minimize the following function.

2 +C]

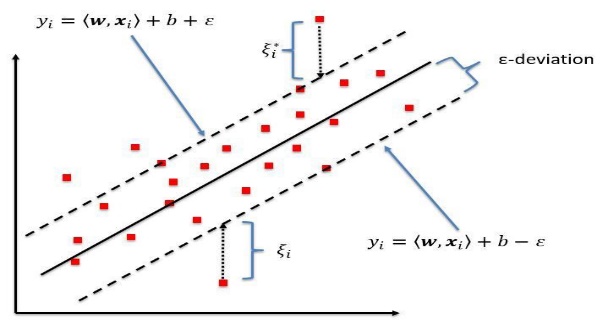
Subject to constraints

y(i)-wx(i)-b ≤ + ξi

wx(i)+ b -y(i)≤ + ξi

The is the margin of tolerance in the above equation. ξi is the misclassification error.

The figure below shows the Support Vector Machine for Regression when there is a slack variable.



**Fig 8**. Support Vector Machine Regression model[9]

For detailed explanation of SVM regression with examples, see references [9] [Full Article](https://www.researchgate.net/publication/320916953_Predicting_Top-of-Atmosphere_Thermal_Radiance_Using_MERRA-2_Atmospheric_Data_with_Deep_Learning),[11] .[Full Article](https://www.researchgate.net/publication/326468246_CAT-PUMA_CME_Arrival_Time_Prediction_Using_Machine_learning_Algorithms)

There are different applications of SVM regression such as Volatile Market Prediction, for detailed explanation of the application ,see reference[10] [Source](http://www.cse.cuhk.edu.hk/~king/PUB/ideal2002.pdf), project control forecasting, for detailed explanation of the application ,see reference [12] [Full Article](https://www.sciencedirect.com/science/article/abs/pii/S0926580514001794), Study on Driving Decision-Making Mechanism of Autonomous Vehicle, for detailed explanation of the application ,see reference [13] [Full Article](https://www.mdpi.com/2076-3417/8/1/13).

# **SVM -OUTLIER DETECTION**

Support Vector Machines are also used for outlier detection. One class Support Vector Machines are used for detecting the outliers. They will separate the normal data from the anomaly data using the hyperplane. The Outlier detection using Support Vector machine has many applications. Some of them are Application of One-Class SVM to Melanoma Prognosis ,for detailed information see reference[14] [Full Article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041295/), Outlier detection for wireless sensor network in harsh environments, ,for detailed information see reference [15] [Full Article](https://www.researchgate.net/profile/Paul_Havinga/publication/221192293_Adaptive_and_Online_One-Class_Support_Vector_Machine-Based_Outlier_Detection_Techniques_for_Wireless_Sensor_Networks/links/5948d6100f7e9b1d9b238105/Adaptive-and-Online-One-Class-Support-Vector-Machine-Based-Outlier-Detection-Techniques-for-Wireless-Sensor-Networks.pdf), Outlier Detection in Breast Cancer Survivability Prediction, ,for detailed information see reference [16] [Full Article](https://link.springer.com/chapter/10.1007/978-3-540-89376-9_10).

# **SVM PROPERTIES[3]**

* SVM is a sparse technique because it requires only few training points called support vectors for future prediction.
* SVM is a maximum margin separator. The Hyperplanes are situated such that they are at maximum distance from the different classes. One of the obejctives of the optimization in SVM is to maximize the distance between the two Hyperplanes. It is essential as the model is training using s sample data,but it has to predict for the unseen data, there might be slight differences in their distributions.
* SVM uses empirical risk minimization and satisfies the duality and convexity requirements.

# **APPLICATIONS**

SVM have applications in different fields such as economics, finance, Management, criminology, Face Recognition, Image Retrieval, Security, fraudulent credit card transactions, Soft Biometrics from Face Images[18] [Full Article](https://link.springer.com/chapter/10.1007/978-3-319-02300-7_8). Some of the important applications are listed below.

## *Face Recognition*

SVM based algorithm is used for Face Recognition. The algorithm is used for both identification and verification. The SVM algorithm creates the decision Hyperplane by treating single individual as a different class. It is trained in a difference space to capture the dissimilarities between two facial images. In identification it is presented with image of unknown person . The algorithm will best estimate the unknown person from the database of known individuals. For verification, the algorithm is presented with an image and the claimed person identity. It will either reject or accept the claim. For detailed explanation of the application ,see reference[17] [Full Article](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=150750)

## *Security*

The unmanned air vehicles popularity has increased exceptionally due to their automatic moving capacity and applications in several domains. This also results in security threat to the security sensitive departments. So, the classification and detection of these amateur drone sounds has gained importance to protect the security of the departments. The SVM algorithm with various kernels is used to classify these sounds accurately. For detailed explanation of the application ,see reference [19] [View Article](https://ieeexplore.ieee.org/document/8616877)

## *Cancer Genomics*

SVM algorithm use` is expanding in cancer genomics. By its powerful classification feature , SVM is facilitating the better understanding of cancer driver genes. Due to its effective performance in classifying high dimensional data and low size data the SVM is used in cancer Genomics. As in the Cancer Genomics the sample size available( positive cases of cancer) is low when compared to the dimensions of the data(gene features). For detailed explanation of the application ,see reference [22] [Full Article](http://cgp.iiarjournals.org/content/15/1/41.full.pdf)

# **ADVANTAGES[24]**

* SVM is used for both Regression as well as Classification tasks. It is stable as small change to data will not affect the decision boundary.
* SVM has good accuracy rate and useful in high Dimension Space.
* Useful for both linear and nonlinear classification(using kernel) of data.
* SVM is effective in cases where the dimensions of sample data are greater than the number of samples.

# **DISADVANTAGES[24]**

* Identifying the subset of useful features which contribute to the model selection is not automatic. So, choosing the subset is the important task in SVM.
* Tuning of several parameters which will affect the classification result is important, as the model success is dependent on the tuning.
* If there are overlapping classes in the data, SVM classifiers are not preferred.

# **CONCLUSION**

The Support Vector Machine algorithm is most powerful for classification. It is very useful in recognizing patterns in complex datasets. SVM Classifier is useful in many applications as it has ability to classify the linearly separable and nonlinearly separable data.

**REFERENCES**

[1]. Support Vector Machines (SVMs) A Brief Overview. Afroz Chekure . In Towards Data-science [Full Article](https://towardsdatascience.com/support-vector-machines-svms-4bcccbd78369)

[2]. Eyo, Edem & Pilario, Karl Ezra & Lao, L. & Falcone, Gioia. (2019). Development of a Real-Time Objective Gas-Liquid Flow Regime Identifier Using Kernel Methods. IEEE Transactions on Cybernetics. 10.1109/ TCYB .2019.2910257.[Full Article](https://www.researchgate.net/publication/332427600_Development_of_a_Real-Time_Objective_Gas-Liquid_Flow_Regime_Identifier_Using_Kernel_Methods)

[3].Yingjie Tian, Yong Shi,Xiaohui Liu, “Recent Advances On Support Vector Machines Research.” In Technological and Economic Development Of Economy 2012 Volume 18(1) [Full Article](https://www.researchgate.net/publication/254321762_Recent_advances_on_support_vector_machines_research)

[4].Xin ZHOU, Ying WU, Bin YANG, “Signal Classification Method Based On Support Vector Machine and High-Order Cumulants. In Wireless Sensor Network,2010 Scientific research [Full Article](https://www.scirp.org/pdf/WSN20100100007_87680380.pdf)

[5] Bagchi, Tapan. (2013). Refining AI Methods for Medical Diagnostics Management. NMIMS Management Review. XXIII. 67 - 90.[Full Article](https://www.researchgate.net/publication/263582095_Refining_AI_Methods_for_Medical_Diagnostics_Management)

[6] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani ,”An Introduction to Statistical Learning with applications in R.” Springer Texts in Statistics. [View Textbook](https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf)

[7] ISBN: Sebastian, R., Vahid M. Python Machine Learning - Second Edition: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. 2nd edition. September 20, 2017

[8] Durgesh K. Srivastsava, Lekha Bhambhu, “Data Classification Using Support Vector Machine.” In Journal of Theoretical and Applied Information Technology 2005-2009.[Full Article](http://www.jatit.org/volumes/research-papers/Vol12No1/1Vol12No1.pdf)

[9] Kleynhans, Tania & Montanaro, Matthew & Gerace, Aaron & Kanan, Christopher. (2017). Predicting Top-of-Atmosphere Thermal Radiance Using MERRA-2 Atmospheric Data with Deep Learning. Remote Sensing. 9. 1133. 10.3390/rs9111133.[Full Article](https://www.researchgate.net/publication/320916953_Predicting_Top-of-Atmosphere_Thermal_Radiance_Using_MERRA-2_Atmospheric_Data_with_Deep_Learning)

[10] Haiqin Yang,LAiwan Chan, and Irwin King ,“Support Vector Machine Regression for Volatile Stock Market Prediction” [Full Article](http://www.cse.cuhk.edu.hk/~king/PUB/ideal2002.pdf)

[11] Liu, Jiajia & Ye, Yudong & Shen, Chenlong & Wang, Yuming & Erdélyi, R.. (2018). CAT-PUMA: CME Arrival Time Prediction Using Machine learning Algorithms .[Full Article](https://www.researchgate.net/publication/326468246_CAT-PUMA_CME_Arrival_Time_Prediction_Using_Machine_learning_Algorithms)

[12] Mathieu Wauters ,Mario Vanhoucke, “Support Vector Machine Regression for project control forecasting.”. In Automation in construction, Science Direct. [Full Article](https://www.sciencedirect.com/science/article/abs/pii/S0926580514001794)

[13] JunYou Zhang,Yaping Liao,Shufeng Wang, Jian Han ,“Study on Driving Decision-Making Mechanism of Autonomous Vehicle Based on an Optimized Support Vector Machine Regression.”. In Journals ,Applied Sciences Volume 8 Issue 1[Full Article](https://www.mdpi.com/2076-3417/8/1/13)

[14] Stephan Dreiseitl, Melanie Osl, Christian Scheibbock, Michael Binder ,“Outlier Detection with One-Class SVMs: An Application to Melanoma Prognosis.” In Journal AMIA Annu Symp Proc. [Full Article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041295/)

[15] Yang Zhang, Nirvana Meratnia, Paul Havinga “Adaptive and Online One-Class Support Vector Machine-based Outlier Detection Techniques for Wireless Sensor Networks.” In 2009 International Conference on Advanced Information Networking and Applications Workshops. [Full Article](https://www.researchgate.net/profile/Paul_Havinga/publication/221192293_Adaptive_and_Online_One-Class_Support_Vector_Machine-Based_Outlier_Detection_Techniques_for_Wireless_Sensor_Networks/links/5948d6100f7e9b1d9b238105/Adaptive-and-Online-One-Class-Support-Vector-Machine-Based-Outlier-Detection-Techniques-for-Wireless-Sensor-Networks.pdf)

# [16] Jaree Thongkam, Guandong Xu, Yanchun Zhang, Fuchun Huang, “Support Vector Machine for Outlier Detection in Breast Cancer Survivability Prediction.” In Asia Pacific Web Conference 2008 Advanced Web and Network Technologies, and Applications. [Full Article](https://link.springer.com/chapter/10.1007/978-3-540-89376-9_10)

[17] P. Jonathon Phillips, “Support Vector Machines Applied to Face Recognition.” In technical report NISTIR 6241, Advances in Neural Information Processing Systems 11, eds. M. J. Kearns, S. A. Solla, and D. A. Cohn, MIT Press, 1999.[Full Article](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=150750)

# [18] Guodong Guo, “ Soft Biometrics from Face Images Using Support Vector Machines.” In: Ma Y., Guo G. Springer, Cham [Full Article](https://link.springer.com/chapter/10.1007/978-3-319-02300-7_8)

[19] M. Z. Anwar, Z. Kaleem and A. Jamalipour, "Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications," in IEEE Transactions on Vehicular Technology*,* vol. 68, no. 3, pp. 2526-2534, March 2019, doi: 10.1109/TVT.2019.2893615.[View Article](https://ieeexplore.ieee.org/document/8616877)

[20] A. Ganapathiraju, J. E. Hamaker and J. Picone, "Applications of support vector machines to speech recognition," in IEEE Transactions on Signal Processing, vol. 52, no. 8, pp. 2348-2355, Aug. 2004, doi: 10.1109/TSP.2004.831018.[Full Article](https://ieeexplore.ieee.org/abstract/document/1315952)

[21] P. Wang, R. Mathieu, J. Ke and H. J. Cai, "Predicting Criminal Recidivism with Support Vector Machine," 2010 International Conference on Management and Service Science, Wuhan, 2010, pp. 1-9, doi: 10.1109/ICMSS.2010.5575352. [Full Article](https://ieeexplore.ieee.org/document/5575352)

[22] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics*. 2018;15(1):41‐51. doi:10.21873/cgp.20063 [Full Article](http://cgp.iiarjournals.org/content/15/1/41.full.pdf)

[23] Abe, Shigeo. (2003). Analysis of Multiclass Support Vector Machines. International Conference on Computational Intelligence for Modelling Control and Automation .[Full Article](https://www.researchgate.net/publication/29687855_Analysis_of_Multiclass_Support_Vector_Machines)

[24] S. Karamizadeh, S. M. Abdullah, M. Halimi, J. Shayan and M. j. Rajabi, "Advantage and drawback of support vector machine functionality," 2014 International Conference on Computer, Communications, and Control Technology (I4CT), Langkawi, 2014, pp. 63-65, doi: 10.1109/I4CT. 2014 .6914146.[Full Article](https://ieeexplore.ieee.org/document/6914146)

[25]. Support Vector Machine Optimization in python [Full Article](https://pythonprogramming.net/svm-optimization-python-machine-learning-tutorial/?completed=/svm-in-python-machine-learning-tutorial/)

[26] Implementing SVM using python scikit learn [Full Article](https://stackabuse.com/implementing-svm-and-kernel-svm-with-pythons-scikit-learn/)