ANALYSIS OF SUPPORT VECTOR MACHINES

By Aswin Manoj

Student No: 7926291

COMP 3190 Research Project

1. Abstract

Machine learning, a subfield of Artificial Intelligence, is used to process, learn, and analyze data to make intelligent decisions based on the data [Black et al, 2022]. This paper briefly talks about machine learning and the algorithms employed for machine learning, specifically Support Vector Machines [Huang et al, 2018]. Linear SVM and Non-Linear SVM are two types of Support Vector Machines [Huang et al, 2006]. Kernel functions are introduced with a focus on its relationship with Non-Linear SVM [Huang et al, 2006]. Support Vector Machines are used in real world applications such as the drug discovery sector and in handling cancer genomic data. While Support Vector Machines are revolutionary, there still exists challenges for Support Vector Machines such as problems with quadratic problem solving [Huang et al, 2006] and underlying complexity [Huang et al, 2018].

2. Introduction

Machine learning is an automated process undertaken by computers which involves learning and finding patterns in data, by using algorithms and other techniques to analyze and produce conclusions [Black et al, 2022]. Selecting the right Machine learning algorithm for a given set of data is necessary for optimal performance and results [Black et al, 2022]. Support Vector Machine is a subfield of supervised machine learning [Huang et al, 2006] and it involves the classification of datasets primarily of a binary type [Heikamp and Bajorath, 2014].

3. An insight into Machine Learning

3.1 Introduction to Machine Learning

Machine Learning has come under attention due to its ability to classify and predict through the availability and access to a large reserve of data [Black et al, 2022]. It achieves this by identifying and learning from various patterns and features in a given data set [Black et al, 2022]. By using past data to discern and model future data, machine learning is used to detect complex patterns from large, jumbled, and complex data sets [Huang et al, 2018]. This kind of ability is used for group classification especially in cluttered and noisy genomic data [Huang et al, 2018]. Being able to predict and assess data intelligently, it is a subfield of Artificial intelligence even though it requires a decent amount of user input [Black et al, 2022]. Supervised Learning Techniques are machine learning techniques which are used to classify and predict data [Black et al, 2022].

3.2 Supervised learning techniques

Supervised learning techniques help the computer predict a plausible output of class for given data points without any observations by providing it data with known observations such as a mapping of inputs and outputs [Badillo et al, 2020]. This enables the computer to create a model from the given known data which it then uses to predict the outputs for a given input without output values [Badillo et al, 2020]. This technique predicts outcomes by learning the interrelation between predictive variables and their known outcomes [Black et al, 2022]. Supervised learning is classified into numerical label and categorical labels depending on the type of output [Badillo et al, 2020]. Numerical labels deal with regression and consequently Support Vector Machine regression is under this label while categorical label deals mainly with classification and under this comes Support Vector Machine without regression [Badillo et al, 2020]. Some examples where this technique is used in are decision trees, random forests and SVM's [Black et al, 2022; Huang et al, 2006].

3.3 Differences between Machine learning and Statistics

Both machine learning and statics work hand in hand [Black et al, 2022]. For example, when machine learning is used in an analysis to predict probabilities, statistical model can then use the outcomes predicted and vice versa [Black et al, 2022]. But Machine Learning can do more than what statistics is capable of [Black et al, 2022]. This can be seen in the k-nearest neighbour algorithm where it used the outcome of entities to predict the outcome for a similar entity without knowing much about the parametric model of that dataset [Black et al, 2022].

4 Introduction to Support Vector Machines (SVM's)

4.1 Introduction

One of the algorithms or learning methods under Machine Learning is Support Vector Machines (SVM's) [Huang et al, 2018]. This supervised learning method is capable of modelling unknown, partially known, highly complex and nonlinear processes as well as "being universal approximators of any multivariate functions to any desired degree of accuracy" [Huang et al, 2006]. This method is mostly used to handle classification or regression problems which was proposed by Cortes in the 1990s [Zhao et al, 2020]. It is primarily used to classify two different classes having class instance values of +1 and -1 through learning from the input data and it does the classification by establishing a line between the two different classes [Efeoğlu and Tuna, 2022]. The line that is produced by the SVM to separate the given dataset is termed as a hyperplane [Huang et al, 2018]. It is located and placed in such a way that it is at the farthest distance from the closest point in the two datasets and these points are called support vectors [Huang et al, 2018]. From figure 1, it is seen that the Support Vector Machine draws a line between the two dataset instances in such a way that the margin is at its maximum value [Efeoğlu and Tuna, 2022]. It is also observed in figure 1 that the class instances closest to the linear line of separation is termed support vectors [Efeoğlu and Tuna, 2022].

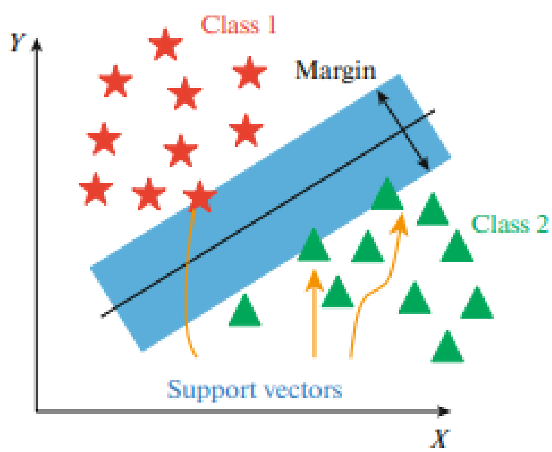

Figure 1, (Efeoğlu and Tuna, 2022)

4.2 <u>Linear SVM</u>

SVM in its basic form is used to classify data that could be separated linearly or to solve binary classification problems [Heikamp and Bajorath, 2014]. A training data for a binary classification is given,

$(x_1, y_1), (x_2, y_2),\ldots,(x_n, y_n),\ x \in R^m,\ y \in \{+1, -1\}$ (Huang et al, 2006).

When considering a linear space, infinite hyperplanes can be constructed that separates or classifies the classes [Heikamp and Bajorath, 2014]. But the goal of SVM is to find a hyperplane that can be used to separate the datasets with the maximum margin possible between the support vectors [Huang et al, 2006].

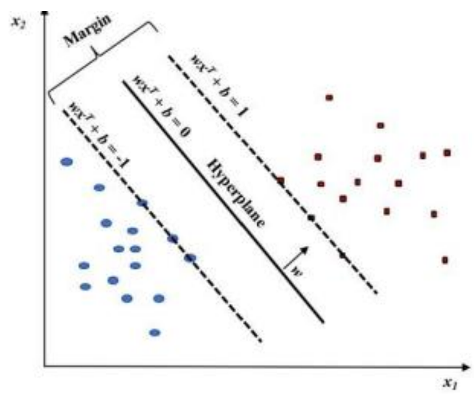

Figure 2, [Huang et al, 2018]

From figure 2, the equation used to find the maximal margin hyperplane is given as,

$wx^T + b = 0$ [Huang et al, 2018].

Here, from the hyperplane equation, w is defined as the weight vector, b is the bias and x is given as the input feature vector [Huang et al, 2018].

"The w and b would satisfy the following inequalities for all elements of the training set:

$wx_i^T + b \geq +1$ if $y_i = 1$

$wx_i^T + b \leq +1$ if $y_i = -1$" [Huang et al, 2018].

4.3 <u>Nonlinear SVM</u>

Till now, the type of pattern recognition task that has been covered is of the simplest type containing classes that don't overlap and where the Support Vector Machine uses a linear hyperplane to classify the datasets with a maximal margin [Huang et al, 2006]. When classes cannot be linearly separated using a linear hyperplane, the Support Vector Machine transforms the original input space into a higher dimensional feature space by using kernel functions that can be used to transform the original nonlinear data by adding more dimensions to turn it into a linear problem [Huang et al, 2018].

After transforming the input space into a feature space, the next task that the Support Vector Machine must do is to create a hyperplane that separates the feature space [Huang et al, 2006]. This can be done in a similar way of how a linear hyperplane with a maximal margin was used to classify and separate the classes in the original input space [Huang et al, 2006].

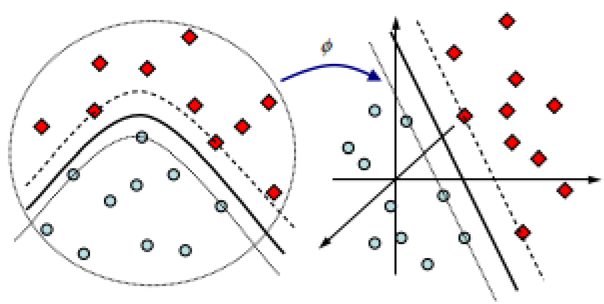

Figure 3, [Chen et al, 2008]

In the given figure 3, a Support Vector Machine performs a linear separation on a space with extra dimensions, where the resulting separation is nonlinear before its transformation [Chen et al, 2008].

A nonlinear SVM is constructed by keeping the main concept in mind which is to "map the input vectors $x_i \in R^m$ into vectors $\Phi(x_i) \in R^s$ of a high dimensional feature space *S* (where Φ represents mapping: $R^m \rightarrow R^s$) and to solve a linear classification problem in this feature space:

$$x \in R^m \rightarrow \Phi(x) = [\Phi_1(x)\ \Phi_2(x), \ldots, \Phi_s(x)]^T \in R^S$$ "

where a fixed function Φ (the mapping) is chosen before hand itself [Huang et al, 2006].

4.4 Kernel Functions

Support Vector Machine which was originally used as a linear classifier for linearly separable data uses kernel functions alternatively so that any type of data set which can be linear or nonlinear can be dealt with by using an appropriate function for each problem [Zhao et al, 2020] and to model a data that is having a higher dimension [Huang et al, 2018]. To find the best kernel function to use is not possible directly and hence the only way to find it is by trial and error [Huang et al, 2018]. The nature of the type of Support Vector Machine used and the problem at hand can give a clue on what kernel can be used [Huang et al, 2018]. One of the methods that can be employed to find the right kernel function for a given problem set is by using cross confirmation on a given set of kernels through a method that contains statistical probing [Huang et al, 2018]. For a kernel function to be valid it must accomplish two conditions where the first one is that it must be symmetric and the second one is that it needs to be positive semi-definite such that a global optimum exists [Heikamp and Bajorath, 2014]. A kernel function becomes symmetric only when $K(x_1, x_2) = K(x_2, x_1)$ [Heikamp and Bajorath, 2014].

"A kernel is a function K such that,

$$K(x_i, x_j) = \Phi^T(x_i)\,\Phi(x_j)\text{" [Huang et al, 2006].}$$

Different types of kernel functions including linear and nonlinear kernel functions are described below:

- A linear kernel function is an uncomplicated function [Huang et al, 2006] where for two observations, the linear function can be used on it by doing a dot product [Huang et al, 2018] given by the equation,

$$K(x_i\ , x_j) = x_i^T x_j \text{ [Huang et al, 2006].}$$

- $K(x\ , x_i) = [(x^T x_i) + 1]^d$ is a polynomial classifier kernel function [Huang et al, 2006] that can be used to distinguish between a curved linear input space and is a general form of a linear kernel function [Efeoğlu and Tuna, 2022] where d is the degree of the polynomial function [Huang et al, 2006].

- Gaussian radian basis function "is an example of a radian basis function kernel" [Efeoğlu and Tuna, 2022] and is given by the equation,

$$K(x_i, x_j) = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right),$$

  "Where σ is a parameter that determines the width of the Gaussian kernel and significantly affects the performance of the kernel" [Efeoğlu and Tuna, 2022].

- Sigmoid kernel function is an example of a hyperbolic tangent kernel [Efeoğlu and Tuna, 2022] and a multilayer perceptron kernel that is conditionally positive definitive [Huang et al, 2006]. If it has a slope α and has a constant δ, it can be given by the equation,

$$K(X_i, X_j) = \tan h(\alpha x_i^T x_j + \delta) \text{ [Efeoğlu and Tuna, 2022].}$$

The advantage of using a kernel function in a Support Vector Machine is to avoid calculating the mapping Φ(x) [Huang et al, 2006]. It is not required to know what the mapping is [Huang et al, 2006] since the kernel function is used directly in the input space [Huang et al, 2018]. A critical thing to be careful of is that the correct kernel function should be chosen for a given Support Vector Machine because if the correct function is not selected, the problem fails even though a correct algorithm is being employed [Efeoğlu and Tuna, 2022]. This leads to unsuccessful results and a dip in the algorithm's performance as compared to its performance with an optimal function [Efeoğlu and Tuna, 2022].

4.5 Challenges and drawbacks of SVM

One major drawback is that Support Vector Machines struggle with solving quadratic programming problems in some situations [Huang et al, 2006]. Classification in its most basic form involves solving these quadratic programming problems [Huang et al, 2006]. Usually, when the size of training data is small, it functions optimally [Huang et al, 2006]. But, when the size of the data set increases, it becomes extremely difficult to solve and requires a large memory for processing [Huang et al, 2006]. Computers nowadays are simply not equipped to handle or store this in their memory [Huang et al, 2006].

Another drawback is that the result of a problem where an SVM is used may depend heavily on the knowledge and expertise of the person using it due to the complexity of the mathematical models and workings that lie beneath the SV functions involved [Huang et al, 2018].

4.6 <u>Applications of SVM's</u>

The application of Support Vector Machines is well known and successful in cancer genomics [Huang et al, 2018]. Cancer genomic data is known for being cluttered [Huang et al, 2018]. SVM is popular with cancer genomic data because of the effective results provided by this learning machine and the ways in which the kernel functions can handle this data by adapting to different circumstances [Huang et al, 2018].

Support Vector Machines has also been used in the drug discovery sector to predict if a compound is active or inactive against single or multi targets [Heikamp and Bajorath, 2014]. It has also been used to predict side effects of different compounds [Heikamp and Bajorath, 2014].

5. <u>Conclusion</u>

This paper has introduced a critical concept in Artificial Intelligence which is Machine Learning [Black et al, 2022]. It is seen that machine learning imitates human behaviour by analysing data and making intelligent conclusions [Badillo et al, 2020]. The paper then shows in depth information on one of the machine learning algorithms which is Support Vector Machines [Huang et al, 2018]. It is seen that Support Vector Machines classify linearly separable data through calculating a hyperplane that separates the data [Heikamp and Bajorath, 2014]. It is also shown handling data which is not linearly separable through kernel functions [Huang et al, 2018]. Support Vector Machines face some challenges such as when the data set increases, it becomes significantly difficult for quadratic problem solving due to the need for large memory storage

[Huang et al, 2006]. The complexity of the mathematics involved in solving problems that use Support Vector Machines is also a drawback (Huang et al, 2018). Analysis of cancer genomic data [Huang et al, 2018] and the prediction of the activity status of a compound against single and multi targets are some of the applications of Support Vector Machines [Heikamp and Bajorath, 2014]. All in all, I believe that Support Vector Machines, in the right hands, is a revolutionary tool that can be used on a multitude of datasets and provide excellent results.

Reference List

[Badillo et al, 2020] Badillo, S. et al. (2020) An Introduction to Machine Learning. Clinical pharmacology and therapeutics. [Online] 107 (4), 871–885.

[Black et al, 2022] Black, J. E. et al. (2022) An introduction to machine learning for classification and prediction. Family practice. [Online]

[Huang et al, 2018] Huang, S. et al. (2018) Applications of support vector machine (SVM) learning in cancer genomics. Cancer genomics & proteomics. [Online] 15 (1), 41–51.

[Huang et al, 2006] Huang, T.-M. et al. (2006) 'Support Vector Machines in Classification and Regression — An Introduction', in Kernel Based Algorithms for Mining Huge Data Sets. [Online]. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 11–60.

[Efeoğlu and Tuna, 2022] Efeoğlu, E. & Tuna, G. (2022) Machine Learning for Predictive Maintenance: Support Vector Machines and Different Kernel Functions. Journal of machinery manufacture and reliability. [Online] 51 (5), 447–456.

[Zhao et al, 2020] Zhao, W. et al. (2020) Applications of Support Vector Machine in Genomic Prediction in Pig and Maize Populations. Frontiers in genetics. [Online] 11598318–598318.

[Heikamp and Bajorath, 2014] Heikamp, K. & Bajorath, J. (2014) Support vector machines for drug discovery. Expert opinion on drug discovery. [Online] 9 (1), 93–104.

[Chen et al, 2008] Chen, S.-H. et al. (2008) A support vector machine approach for detecting gene-gene interaction. Genetic epidemiology. [Online] 32 (2), 152–167.