

PAPER • OPEN ACCESS

## A Brief Conceptual View on Classification Using Support Vector Machine

To cite this article: Nur Fatin Liyana Mohd Rosely *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **551** 012035

View the [article online](#) for updates and enhancements.



**The Electrochemical Society**  
Advancing solid state & electrochemical science & technology

The ECS is seeking candidates to serve as the  
**Founding Editor-in-Chief (EIC) of ECS Sensors Plus,**  
a journal in the process of being launched in 2021

The goal of ECS Sensors Plus, as a one-stop shop journal for sensors, is to advance the fundamental science and understanding of sensors and detection technologies for efficient monitoring and control of industrial processes and the environment, and improving quality of life and human health.

*Nomination submission begins: May 18, 2021*



Nominate now!

# A Brief Conceptual View on Classification Using Support Vector Machine

*Nur Fatin Liyana Mohd Rosely*<sup>1</sup>, *Azlan Mohd Zain*<sup>1</sup>, *Yusliza Yusoff*<sup>1</sup> and *NorFadzlan Yusup*<sup>2</sup>

<sup>1</sup>School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor

<sup>2</sup>Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak

Email: yanarosely@gmail.com

**Abstract.** Classifier is an important element for a classification process to classify features or samples of a dataset successfully. The existence of classifier is important, especially involved multiple class dataset with huge number of features or samples. Support Vector Machine (SVM) classifier catches more attention over other classifiers as it is widely used in various fields. Therefore, this paper aims to give a brief conceptual view of SVM in classification. The important elements and principle of SVM are also described. Along with that an appropriate evaluation and validation for SVM performance are also included. As a result, SVM concept in solving real world classification problem is identified for future reference.

## 1. Introduction

Classification is a supervised learning process that the input space is mapped into predefined classes and it can be done through supervised learning algorithm. In classification case, the learner act as function to map a vector that corresponding to its class using input-output examples. These examples consist of an input object and a desired output value. In order to accomplish this, a help from classifier is needed. A classifier is an abstract meta-class classification that describes common behavior or structural features for a set of instances. Figure 1 shows what classifier is.



**Figure 1.** Classification Conceptual View.

Based on the Figure 1, the input represents the data from an individual. It can be an expression profile of a tumor, meteorological measurement, digitized fingerprint scans or any type of data that technically consist several of the class. The output represents the class of the individual. In order to produce accurate classification output, a classifier is needed. There are many alternative classifiers such as k-nearest neighbor (KNN) algorithm, artificial neural network (ANN) and support vector machine (SVM). Among them, the SVM has a bias among the researchers [1, 2, 3, 4, 5, 6, 7, 8]. Further discussion about support vector machine is discussed in Section 2. Section 2 is organized into

three subtopic of SVM. Performance validation offered to evaluate the SVM is viewed in Section 3. Section 4 briefly review a few related studies to SVM by previous researchers.

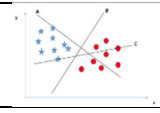
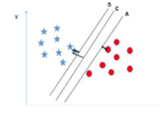
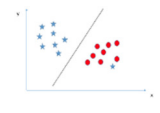

## 2. Support Vector Machine

Support vector machine is the state of the art classification techniques [9]. SVM is known as classification techniques with computational advantages over their contend showers [10]. SVM promise generalization performance in classification and it depends on the principle of structural risk minimization [4, 6, 11].

### 2.1. SVM Elements and Principle

In SVM there are two main elements; a) support vectors represent the points in datasets b) decision boundary represents hyperplane. The distance between all these points and the line is the maximum possible among the possible decision boundaries that lead to produce the most optimal one. The line that successfully segregates the points into classes better among the lines is defined as the optimal. Table 1 shows possible scenarios in SVM that lead to different solutions to obtain most optimal hyperplane.

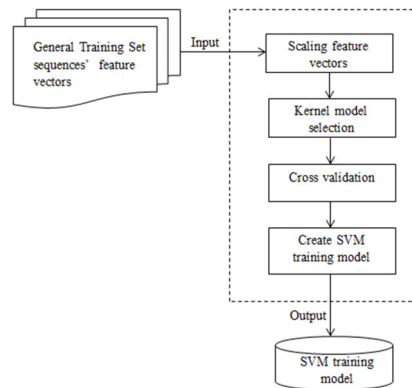
**Table 1.** Scenario and Solution Using SVM.

Scenario	Situation and Solution
	Situation: To identify the right hyper plane to classify star and circle. Solution: Follow thumb rule- Select the hyperplane which segregates the two classes better. Hyperplane B is selected.
	Situation: Three hyper-planes (A, B, C) to segregate the classes (Star and Circle) well. Solution: Maximize the distance between nearest data point (either class) and hyperplane. Based on figure, hyperplane C is high as compared to both A and B. Then, hyperplane C is selected.
	Situation: Unable to segregate the two classes using a straight line. One of the stars lies in the territory of circle as an outlier. Solution: Ignore outlier and find the hyperplane that has maximum margin. Hence, the hyperplane as shown.
	Situation: Cannot have linear hyperplane between two classes. Solution: Using kernel trick.

Based on Table 1, Scenario 1 till 3 involves linear separable data and considered as linear classification. Meanwhile, Scenario 4 is unable to be classified using linear decision boundary. In this situation, kernel plays an important role. Technically, the kernel takes two inputs, spits out them and grouped how alike they are. There are two types of kernel, single and multiple. Referred to [12] single kernel needs to choose proper parameters and multiple kernel is required to search either linear or non-linear combination of predefined base kernels by maximizing the margin maximization. Therefore, multiple kernel provides more flexibility in solving similarities of dataset compared to single kernel.

### 2.2. General Framework

Generally, SVM performance is depending on value of parameters in the training phase. Besides kernel function, regularization parameter ( $c$ ) and gamma ( $g$ ) are the important parameter in SVM. The  $c$  parameter determines the trade-off cost between the training error and the complexity of the model. Meanwhile,  $g$  parameter defines the non-linear mapping from the input space to some high dimensional feature space. Figure 2 reflects the general framework of SVM.



**Figure 2.** General Framework of SVM.

Relate in solving problem using the SVM, the general equation is defined as equation (1), equation (2) represents the hyperplane.

$$W(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle \tag{1}$$

where,

$$\alpha_i \geq 0, i = 1, \dots, m \quad \sum_{i=1}^m \alpha_i y_i = 0$$

$$f(x) = \langle w \cdot x \rangle + b = \sum \alpha_i y_i \langle x_i \cdot x \rangle \quad \text{with} \quad w = \sum \alpha_i y_i x_i \tag{2}$$

where,

weight (w) represented only by data points

$$\text{kernel function } K(x_1, x_2) = \langle \phi(x_i) \cdot \phi(x_j) \rangle$$

### 2.3. SVM for Classification

The kernel implicitly consists of nonlinear transformation reflect SVM has the ability to model nonlinear relationship. SVM do not suffer from deficiency therefore SVM superior performance in classifying high dimension and sparse data [13]. Besides, SVM has advantages compared to others classifier because the value of c and g is chosen optimally [14]. As an example, neural network (NN) traps to local minima and reflects NN not robust over different samples.

Sahran et al. [15] proposed Support Vector Machine-Recursive Feature Elimination (Absolute Cosine) (SVM-RFE(AC)) to overcome overfitting issue due to the high dimensional texture of individual tissue components. SVM-RFE(AC) is integrated embedded procedure of SVM-RFE with filter method of AC. This integration helped prevent redundancy in selected features of SV-RFE and a un-optimised classifier in the AC. The results show SVM-RFE(AC) superior to SVM and SVM-RFE. It has proven the proposed SVM-RFE(AC) able to identify of the most crucial texture of each tissue component.

Qi et al. [12] applied SVM as classifier where the process involves feature selection and multiple kernel boosting frameworks based on PSO. Then it was tested through experiment that using HyperSpectral Image (HSI) datasets as a benchmark. The results reflect the proposed approach can achieve better accuracy and efficiency than state of the art methods. Differ to Rajamohana and Umamaheswari [16] used SVM to classify the performance of proposed method which is improved binary particle swarm optimization and shuffled frog leaping algorithm. The main of objective of the study is to avoid spammers manipulate the product or service reviews in their favor either devalue or promote purposed.

### 3. Performance Validation

Generally, cross validation and performance measurements are used to evaluate and validate how good a classification model. As SVM is one of classification models, these also applied to SVM. Normally, cross validation involved training and testing data and performance measurement relate to evaluation of the final result of model performance.

#### 3.1. Cross Validation

Cross validation is divided into two types, exhaustive cross validation and non-exhaustive. Exhaustive cross validation guarantee learn and test on all possible ways to divide the original sample into training and a validation set. Leave-one-out and leave-p-out are categorized into this type. Meanwhile, non-exhaustive cross validation does not compute all ways of spitting the dataset. Holdout method and k-fold cross validation are methods of non-exhaustive cross validation. Table 2 summarizes the cross validation methods.

**Table 2.** Summary of cross validation methods.

Method	Description
Leave-One-Out	1. Specific case of k-fold cross validation with K=N (number of data) 2. High computation time, however a good practice to validate data.
Leave-p-Out	1. In N data, N-P is taken in one iteration and the remaining are used as validation. 2. The high value of p produces more exhaustive.
Holdout Method	1. Simplest types of cross validation and using the training set only. 2. Short time to compute, but have high variance for evaluation.
k-Fold Cross Validation	1. A single subsample is reserved as validation data for testing part and remaining k-1 are treated as training data. 2. Less matter how the data gets divided and selection not bias.

#### 3.2 Performance Measurement

Basically, the performance measurements are based on accuracy (Ac), sensitivity (R), precision (P), specificity (Sp) and F-Measure (F1).. The Ac is a metric proportion of identifying the class label of the new and invisible data. R (also known as recall) and Sp is used to evaluate the obtain feature subset. The R represents the positive and Sp represents the negative. They are divided into true positive (TP), true negative (TN), false negative (FN) and false positive (FP). True represent the correct classification and false is the incorrect classification. The P is used to evaluate the feature subset to all available features. The F1 is the combination of P and R and it known as harmonic mean. The formula of these measurements as stated as follows

$$\text{Accuracy (Ac)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Specificity (Sp)} = \frac{TN}{TN + FP} \tag{4}$$

$$\text{Sensitivity (R)} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Precision (P)} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{F-Measure (F1)} = \frac{2TP}{2TP + FP + FN} \tag{7}$$

### 4. Related Studies

Based on the above description, this section will discuss several recent studies related to SVM. Arumugam an Jose [4] applied SVM with filter Decision tree to solve problem related to computational time of labelled training data. The modification taken by recuperate the data points that

are near to the decision boundaries. This study uses 10 different dataset from <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html> to evaluate the proposed method. They are leukemia, Duke breast cancer, Wisconsin Prognostic Breast (WPBC), Breast Cancer Wisconsin (Diagnostic) Data Set (WDBC), Gisette, IJCNN1, epsilon, susy, KDDcup 2010 and Higgs. 10 fold cross validation is used as performance validation.

Bouazza et al [7] used SVM as a classifier to perform studies of feature selection that are applied to five different microarray datasets. The studies focused on the notion of maximum margin and kernel unction of SVM. The TP, TN, FP, FN and accuracy are used as the performance measurement. Basically, the accuracy of SVM is compared to others classifier such as K Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Decision Tree for Classification (DTC) and Naïve Bayes (NB).

Dash [6] used SVM with Hybrid Harmony Search and Pareto optimization in order to overcome the difficulties in the selection of the most significant features. SVM proof not easily get stuck at local optimum and good approach for binary classification problem. The result through accuracy, specificity and sensitivity shows SVM performed better than other classifier (KNN, NB and ANN).

## 5. Conclusion

From the Section 2, the definition of the SVM concept for classification is described conceptually. The main concept and important element in SVM are explained. Through this, the readers able to notify the basic needed for a standard SVM. Meanwhile, Section 3 provides the available evaluation and validation methods for a SVM model. As conclusion, by having better understanding of the basic conceptual of SVM, the development to enhance and modify the SVM will be easier.

## Acknowledgements

Special appreciation to reviewers for useful suggestions. The authors greatly acknowledge the Research Management Centre, UTM and Ministry of Higher Education Malaysia (MOHE) for financial support through the UTM Zamalah. Also the Research University Grant (RUG) for financial support through grant vot. No. Q.J130000.3551.05G41.

## References

- [1] Jiang H *et al* 2017 *Knowl Based Syst.* **138** 79.
- [2] Jain I *et al* 2018 *App. Soft Comput.* **62** 203.
- [3] Apolloni J *et al* 2016 *App. Soft Comput.* **38** 922.
- [4] Arumugam P *et al* 2018 *Materials Today: Proceedings* **5** 1.
- [5] Ahila R *et al* 2015 *App. Soft Comput.* **32** 23.
- [6] Dash R 2018 *J King Saud Univ. Comp. & Info. Sci.* DOI.
- [7] Bouazza S H 2018 *Procedia Comp. Sci.* **127** C.
- [8] Tuo Y *et al* 2018 *Molecular Med. Rpt.* **17** 3.
- [9] Vapnik V *et al* 1997 *Neural Info. Proc. Sys.* **9** 281.
- [10] Cristianini N *et al* 1999 *Adv. Neural Info. Proc. Sys.* **11** 204.
- [11] Mishra S *et al* 2015 *Int. J Mod. Sci.* **1** 2.
- [12] Qi C *et al* 2017 *Neurocomputing* **220** 181.
- [13] Kusairi R M *et al* 2017 *Int. J Adv. Sci. Eng. Infor. Tech.* **7** 4-2.
- [14] Ren Y *et al* 2010 *JCP* **5** 8.
- [15] Sahran S *et al* 2018 *AI in Med.* **87** 78.
- [16] Rajamohana S P *et al* 2018 *Comp. & Elec. Eng.* **67** 497.