

Brain Cancer Multiclass Classification Using Support Vector Machines Method

Zuherman Rustam¹; VinezhaPanca²

^{1,2}Department of Mathematics, University of Indonesia, Depok 16424, Indonesia

E-mail: rustam@ui.ac.id¹; vinezha.panca@sci.ui.ac.id²

ABSTRACT

This paper will explain about multiclass classification for Brain Cancer using Support Vector Machines (SVM) method. The classifications consist of four disease classes (Medulloblastoma, Malignant Glioblastoma, Primitive Neuroectodermal Tumours and Atypical Teratoid Rhabdoid Tumour) and one non brain cancer class. There are two approaches to be used by applying SVM : one-versus-one and one-versus-rest. For each approaches, there are four SVM kernels used : linear, polynomial, sigmoid, and radial basis function. Based on the observation results, the optimum accuracy for one-versus-rest approach is 84.39% and the optimum accuracy for one-versus-one approach is 78.69%.

Keywords: brain cancer, multiclass classification problem, Support Vector Machines, one-versus-rest, one-versus-one.

1. INTRODUCTION

The brain cancer is an anomalous body cell accretion that occurs in the center nerves system. It is the most common type of cancer that occurs to people below 40 years old, as well as the largest cause of mortality [1]. However, it lacks adequate research compared to other classes of cancer. By applying the machine learning methods, designed and developed by mathematicians, brain cancer can be easily detected.

Generally, the machine learning method can be utilized in detecting and classifying cancerous diseases based on its attributes or characteristics. Furthermore, its characteristic pattern is obtained from training data comprising of class samples, and generalized to form the model for predicting class from the new sample not involved in the learning process. This kind of learning is known as supervised learning.

Support Vector Machines (SVM) is a method used to analyze the classification problem, by determining the hyperplane in order to maximize the margin where its distance is closest to the data point from each class using the SVM concept [2]. Initially, SVM is used to solve binary or two-class problems, but it can also be used to solve multiclass problem by dividing it into two classes problems using the one-versus-rest and one-versus-one techniques [3].

This study calculates the accuracy of the SVM method that is applied in brain cancer multiclass data, using one-versus-rest or one-versus-one approach. The research outcome would be used by doctors or medical expertise in the field of brain cancer to predict its kind and quickly recommend medication. Literature study on SVM for multiclass and simulation program using R programming language were used in conducting this research.

2. BRAIN CANCER MULTICLASS CLASSIFICATION BY SUPPORT VECTOR MACHINE METHOD

2.1 Support Vector Machine

Support Vector Machine (SVM) is a method used to determine the hyperplane that maximizes the margin of the distance closest to the data of each class. The error value is smaller at the generalization when the margin is bigger. The hyperplane formula is denoted by $f(x) = w \cdot x + b$ [5] in which w is the weight parameter and b the bias parameter. Generally, the mathematic model of Support Vector Machine primal optimization is [5]:

$$\min \frac{1}{2} \|w\|^2$$

$$s.t. \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) \geq 1 \quad i = 1, 2, \dots, N \quad (1)$$

The error tolerance factor of Support Vector Machine is written as follows:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

$$s.t. \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) \geq 1 - \xi_i \quad i = 1, 2, \dots, N$$

$$\xi_i \geq 0 \quad i = 1, 2, \dots, N \quad (2)$$

, $C > 0$ is the regulation parameter used to minimize the classification error, and maximize the hyperplane margin, while ξ_i is a variable slack. The value criteria are as follows [5]:

- If $\xi_i = 0$, i^{th} data is at the margin or at the right class side.
- If $0 < \xi_i \leq 1$, i^{th} data is in the margin area and on the right class side.
- If $\xi_i > 1$, i^{th} data is at the wrong class side with classification error.

To complete the SVM problem, the primal was transformed to a dual optimization problem by using Lagrange multiplier. In addition, the formed problem model is written as follows [5]:

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^N \alpha_i$$

$$s.t. \quad \sum_{i=1}^N y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N \quad (3)$$

In which α_i is Lagrange multiplier.

After obtaining the α_i by numerical approach [2], the weight vector and bias parameter value are stated as follows: [2]

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \quad (4)$$

$$\mathbf{b} = \frac{1}{N_S} \sum_{i \in S} (y_i - \sum_{m \in S} \alpha_m y_m \mathbf{x}_m) \quad (5)$$

Therefore, the optimum hyperplane equality is formed.

2.2 One-Versus-Rest

In one-versus-rest approach, for a multiclass problem with k classes, k two-class problems are formed. In each two-class problem, one class (namely class 1) consists all data of a certain class and the other class consists data of all other classes (namely class -1). For example, if there are 3 classes called A, B, and C, the two-class problems will be A and not A, B and not B, C and not C. The final class is obtained by choosing the class with biggest classification score among all two-class classification problems. This class decision method is known as winner-takes-all [6].

2.3 One-Versus-One

In the one-versus-one approach, for a multiclass problem with k classes, there are two-class problems formed. Each two-class problem consists of possible pairing of classes. For example, if there are 3 classes called A, B, and C, the two-class problems will be A and B, B and C, A and C. The final class is obtained by choosing the class which is the most often to appear as classification results among the classification problems. This class decision method is called max-wins [6].

2.4 K-Fold Cross Validation

K-fold cross validation is a method used to conduct validation, in which the data is divided into 1 testing part and $k-1$ training parts. This accuracy is founded after training which creates the second part as the testing data accuracy. The training is done using all of the training parts, and the accuracy is obtained by testing the model using the test data. Then, do the same procedure, this time using the second part of data as testing part and the rest as training parts. After each part has ever become the test data, the overall accuracy is the mean from the accuracy values resulted. [4]

2.5 Kernel

For the mapping $\phi(\mathbf{x})$ in feature space, kernel is defined as [2] :

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\Phi}(\mathbf{x})^T \boldsymbol{\Phi}(\mathbf{x}) \quad (6)$$

Some kernel at SVM [2]:

Linear : $\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \quad (7)$

Polynomial :

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^d \quad (8)$$

Linear :

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\mathbf{a} \mathbf{x}_i^T \mathbf{x}_j + \mathbf{b}) \quad (9)$$

Radial Basis Function :

$$k(x_i, x_j) = \frac{-\|x_i - x_j\|^2}{2\sigma^2} \quad (10)$$

3. RESULTS

The sources used as the data in this research were obtained from <http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi>. It contains 42 samples with 4 brain cancer type and 1 normal class. Each of the type is Medulloblastoma (10 samples), Malignant Glioblastoma (10 samples), Primitive Neuroectodermal Tumours (8 samples), Atypical Teratoid Rhabdoid Tumour (10 samples), and Normal (4 samples). There are 7129 kinds of genes or features used in these data.

The SVM was implemented by adapting the library e1071 at R with four simulations (each simulation uses different kernel) conducted using one-versus-rest and one-versus-one approaches. Furthermore, the validation was conducted using the k-fold cross with $k=3$. Accuracy and running time are the two aspects to be evaluated. Accuracy shows the number of samples in testing data that was correctly predicted. Running time shows the computation duration required to hold learning and prediction. The following are the simulation results: Linear Kernel

The first simulation uses Linear Kernel with various C parameter values, and the optimum accuracy result is achieved when $C=1$.

Table 1. The Simulation with Linear Kernel by One-Versus-Rest Approach

| | The parts of data testing | | | Mean |
|-----------------------|---------------------------|---------|---------|----------|
| | 1 | 2 | 3 | |
| Accuracy (%) | 76.471 | 83.33 | 92.31 | 84.3903 |
| Running Time (second) | 7.44603 | 7.92604 | 7.52503 | 7.632363 |

Table 1 shows the Linear Kernel simulation which uses one-versus-rest approach resulting in a mean accuracy of 84.3903% with mean of running time of 7.632363 seconds.

Table 2. The Simulation with Linear Kernel by One-Versus-One Approach

| | The parts of data testing | | | Mean |
|-----------------------|---------------------------|---------|---------|----------|
| | 1 | 2 | 3 | |
| Accuracy (%) | 76.471 | 83.33 | 76.923 | 76.6968 |
| Running Time (second) | 9.62123 | 9.35023 | 9.83743 | 9.602966 |

Table 2 shows the linear kernel simulation using one-versus-one approach resulting in an accuracy of 76.6968% with mean running time of 9.602966 seconds.

3.1 Polynomial Kernel

The second simulation uses polynomial kernel with $C=1000$. and degree $d=2$.

Table 3. The Simulation with Polynomial Kernel by One-Versus-Rest Approach

| | The parts of data testing | | | Mean |
|------------------------|---------------------------|---------|---------|----------|
| | 1 | 2 | 3 | |
| Accuracy (%) | 76.47 | 83.33 | 84.62 | 81.4733 |
| Running Time (seconds) | 9.34254 | 10.8696 | 9.84256 | 10.01824 |

Table 3 shows the polynomial kernel simulation using one-versus-rest approach which results in a mean accuracy and running time of 81.4733% and 10.01824 seconds respectively.

Table 4. The Simulation with Polynomial Kernel Using One-Versus-One Approach

| | The parts of data testing | | | Mean |
|------------------------|---------------------------|---------|----------|----------|
| | 1 | 2 | 3 | |
| Accuracy (%) | 76.47 | 75 | 84.62 | 78.6967 |
| Running Time (seconds) | 13.8958 | 13.8608 | 13.44777 | 13.73478 |

Table 4 shows the polynomial kernel simulation using

one-versus-one approach which results in a mean accuracy of 78.6967% a running time of 13.73478 seconds.

3.2 Sigmoid Kernel

The third simulation uses sigmoid kernel with $C=1000$, $\gamma=1/7129$, ($\gamma=1/\text{data dimension}$), and $\gamma=0.1$.

Table 5. The Simulation with Sigmoid Kernel by One-Versus-Rest Approach

| | The parts of data testing | | | Mean |
|------------------------|---------------------------|---------|---------|----------|
| | 1 | 2 | 3 | |
| Accuracy (%) | 70.59 | 66.67 | 92.31 | 76.5233 |
| Running Time (seconds) | 10.1176 | 10.1196 | 10.8606 | 10.36593 |

Table 5 shows the kernel sigmoid simulation using one-versus-rest approach resulting in a mean accuracy of 76.5233% and running time of 10.36593 seconds.

Table 6. The Simulation with Kernel Sigmoid by One-Versus-One Approach

| | The parts of data testing | | | Mean |
|------------------------|---------------------------|---------|---------|----------|
| | 1 | 2 | 3 | |
| Accuracy (%) | 76.47 | 75 | 84.62 | 78.6967 |
| Running Time (seconds) | 14.4368 | 13.8278 | 14.5358 | 14.26682 |

Table 6 shows the kernel sigmoid simulation using one-versus-one approach that results in a mean accuracy of 78.6967% with running time 14.26682 seconds.

3.3 Radial Basis Function Kernel

The fourth simulation uses kernel radial with $C=1000$ and $\gamma=0,0000001$, value $\gamma = \frac{1}{2\sigma^2}$.

Table 7. The Simulation with Kernel Radial by One-Versus-Rest Approach

| | The parts of data testing | | | Mean |
|--------------|---------------------------|---------|---------|----------|
| | 1 | 2 | 3 | |
| Accuracy | 76.471 | 83.33 | 84.62 | 81.4737 |
| Running Time | 9.49843 | 9.66683 | 9.91215 | 9.692472 |

Table 7 shows the kernel radial simulation using one-versus-rest approach resulting in a mean accuracy of 81.4737% and a running time of 9.692472 seconds.

Table 8. The Simulation with Kernel Radial by one-versus-one Approach

| | The parts of data testing | | | Mean |
|-----------------------|---------------------------|---------|---------|----------|
| | 1 | 2 | 3 | |
| Accuracy (%) | 76.47 | 75 | 84.62 | 78.6967 |
| Running Time (second) | 16.7715 | 13.1828 | 14.6628 | 14.87239 |

Table 8 shows the kernel radial simulation using one-versus-one approach which results in a mean accuracy of 78.6967% and a running time of 14.87239 seconds.

3.4 Performance Comparison

The Fig.1 show the comparison of the simulations results for one-versus-rest approach.

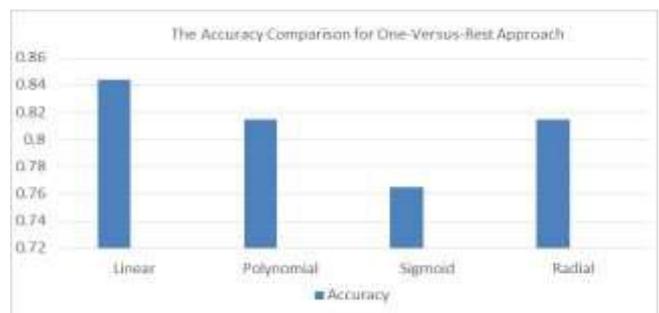


Fig. 1. The Accuracy Comparison for One-Versus-Rest Approach

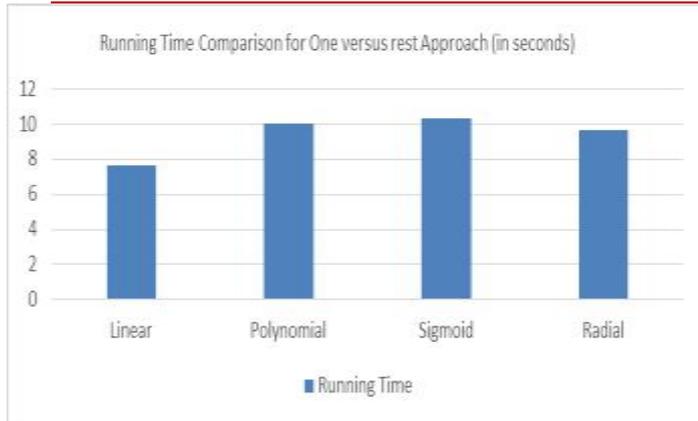
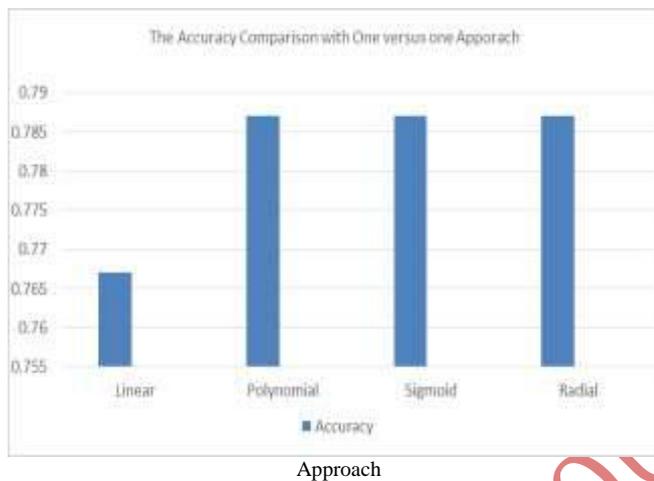


Fig.2 . The Running Time Comparison for One-Versus-Rest



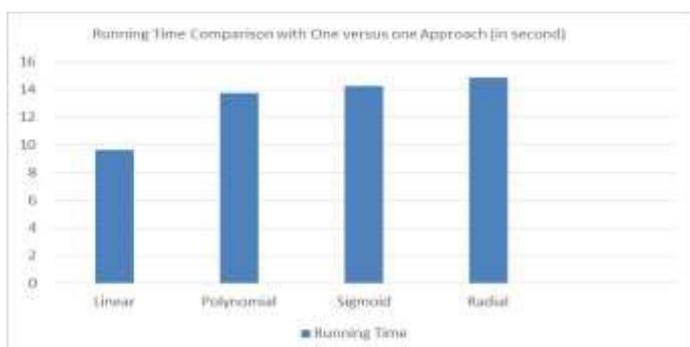
Approach

Based on fig. 1, the highest accuracy for the one-versus-rest approach is achieved using linear kernel (84.3903%), and the lowest accuracy is achieved using sigmoid kernel (76.5233%). Based on Fig 2, fastest running time is obtained using the linear kernel (7.632363 seconds) and the slowest running time is achieved using sigmoid kernel (10.36593 seconds).

The Fig 3 show the comparison of the simulations results for one-versus-one approach.

Fig. 3. The Accuracy Comparison for One-Versus-One Approach

Fig.4. The Running Time Comparison for One-Versus-One Approach



Based on fig. 2, the highest accuracy for the one-versus-one approach is achieved using polynomial, sigmoid, and radial kernel (78.6967%) ,and the lowest accuracy is achieved using linear kernel (76.6968%). Based on fig.4 , the fastest running time is obtained using the linear kernel (9.602966 seconds) and the slowest running time is achieved using radial kernel (14.87239 seconds).

4. CONCLUSION

In this paper, the two approaches used in determining the brain cancer multiclass problem using the Support Vector Machine method are exercised : one-versus-rest and one-versus-one. In addition, the kernels used werelinear, polynomial, sigmoid, and radial. Based on the simulation, the highest accuracy mean is obtained in the approach of one-versus-rest by using linear kernel (84,39%). Furthermore, the simulation by using the three non-linear kernels (polynomial, sigmoid, and radial) in one-versus-one approach results to same mean accuracy (78,69%). Future studies should attempt to determine the optimum parameter value at each kernel to optimize the accuracy of SVM classifier, so that it will be more reliable to help doctors to detect and classify brain cancer. Other possibility is to apply these framework and methodologies in other classification problems, with two or multi classes.

5. REFERENCES

- [1] <https://www.curebraincancer.org.au/page/8/facts-stats> accessed on 12 January 2017 10.51 pm
- [2] Bishop, C.M., 2006, Pattern Recognition and Machine Learning, New York: Springer
- [3] Hsu, C.W. & Lin, C.J., 2002, A Comparison of Methods for Multiclass Support Vector Machines, IEEE Transactions on Neural Networks,13(2), pp.415-425
- [4] Refaeilzadeh, P., Tang,L., dan Liu, H. (2009). Cross Validation, in Ozsu,M.T., andLiu,L (Ed). Encyclopedia of Database Systems, Berlin : Springer
- [5] Panca, V.& Rustam, Z., 2016, Application of Machine Learning on Brain Cancer Multiclass Classification, submitted on The 2nd International Symposium on Current Progress in Mathematics

and Sciences 2016 (ISCPMS), in press

Duan, K.B. & Keerthi, S.S,2005, Which Is the

Best Multiclass SVM Method? An Empirical Study, in N.C. Oza et al., MCS 2005, LNCS 3541, pp. 278–285,

ijournals