A comparative study between support vector machine and support vector data description in bank telemarketing

Han Gao*; Pei Shan Fam; Heng Chin Low

School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia
e-mail: alyssagaohan@gmail.com; fpeishan@gmail.com; lowhengchin@gmail.com

Abstract

Machine learning (ML) is playing more and more important role in the banking system. A supervised ML technique, support vector machine, and an unsupervised ML technique, support vector data description, are applied in the bank telemarketing area to predict that who possibly will buy the term deposit. The results show that SVM outperforms the SVDD based on the overall accuracy and AUC values. The overall accuracy for SVM model attains to 97.30% while the accuracy for SVDD model is a bit higher than random guessing. The AUC values are 0.9867 and 0.8565 for SVM and SVDD model, respectively. SVM model is more suitable for binary classification problem compared to SVDD.

Keywords: supervised learning; unsupervised learning; binary classification; optimization problem

1. Introduction

Due to the highly competitive environment, direct marketing is getting more and more popular and useful in modern market (Elsalamony, 2014). Telemarketing, as a representative direct marketing tool, is an effective and convenient way to sell products and services as well as growing business in several industries, such as medicine, insurance and finance. It began in the early 1990s along with the development of predictive dialer technology (Hurst, 2008). In banking sector, telemarketing plays an irreplaceable role in marketing. In recent decades, more and more machine learning (ML) techniques are applied in this area (Alon et al., 2001; Huang et al., 2007; Mazhar et al., 2007; Moro et al., 2015).

In Moro et al.'s (2014) research, they compared four ML models, namely, logistic regression (LR), decision tree (DT), artificial neural network (ANN) as well as support vector machine (SVM). All the four models are supervised machine learning approaches and no unsupervised models are applied in their research. The biggest difference between supervised and unsupervised learning is that labelled data are used in supervised learning while unlabeled data are used in unsupervised learning. Recently, unsupervised machine learning is gaining more and more popularity in various areas. Therefore, we aim to compare the prediction performance of SVM, a supervised learning, and support vector data description (SVDD), an unsupervised learning in the bank telemarketing area.

Rest of the study is organized as follows: Section 2 presents the methodology used in this research. Section 2.1 and 2.2 discuss SVM and SVDD model, respectively. Section 3 describes the results and discussions from two models mainly based on AUC values. A brief conclusion is displayed in the last section.

2. Methodology

A supervised ML, SVM, and an unsupervised ML, SVDD, are applied to compare the prediction performance in the bank telemarketing based on the dataset collected from a Portuguese banking institution. The mathematical expressions as well as the figures of the two model are displayed in Section 2.1 and 2.2, respectively.

2.1 Support Vector Machine (SVM)

Support vector machine (SVM) is a popular and powerful discriminative classifier in supervised ML area based on the statistical learning theory and the structural risk minimization principle (Vapnik, 1998). There are two fundamental principles of SVM models when dealing with non-linear classification issues, which are the optimal hyperplane and the selection of kernel function (Yao et al., 2008). Figure 1 displays a typical linear and non-linear SVM.

Given the training dataset $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}, y_i \in \{-1, +1\}$ with two labeled classes in the sample space χ, the objective of a support vector classifier is to define an optimal hyperplane to separate the two classes. The mathematical expression of the hyperplane can be shown as:

$$\boldsymbol{w}^T \boldsymbol{x} + b = 0 \tag{1}$$

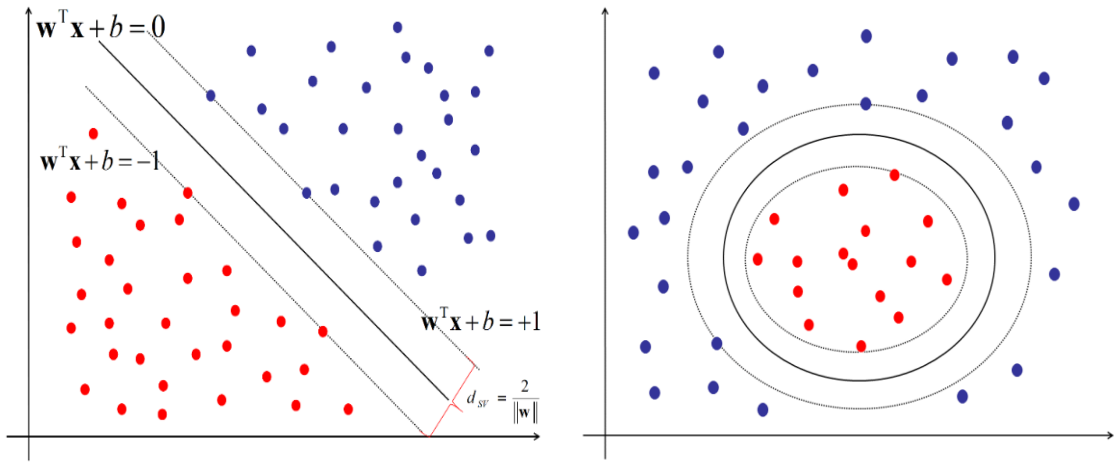
where $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ denotes the normal vector which determines the direction of the hyperplane and b is the displacement which determines the distance from the origin to the hyperplane. Due to the frequent reference, the hyperplane can be denoted as $(\boldsymbol{w}, b)$. According to the distance from point to plane formula, the distance from a sample to the hyperplane can be denoted as:

$$d = \frac{|\boldsymbol{w}^T \boldsymbol{x} + b|}{\|\boldsymbol{w}\|} \tag{2}$$

The objective of SVM is to find a hyperplane with the maximum margin, which can be expressed as an optimization problem shown as:

$$\begin{gathered} \text{minimize: } \frac{1}{2}\|\boldsymbol{w}\|^2 \\ \text{subject to: } y_i(\boldsymbol{w}^T \boldsymbol{x} + b) \geq 1 \;,\; i = 1,2,\ldots,n \end{gathered} \tag{3}$$

In order to save space, more detailed mathematical expressions are not provided here, the interested reader can refer to Wang (2005) and Noble (2006).


$\mathbf{w}^T\mathbf{x}+b=0$
$\mathbf{w}^T\mathbf{x}+b=-1$
$\mathbf{w}^T\mathbf{x}+b=+1$
$d_{SV} = \frac{2}{\|\mathbf{w}\|}$

(a) (b)

Figure 1. A typical (a) linear and (b) non-linear SVM.

2.2 Support Vector Data Description (SVDD)

SVDD, proposed by Tax & Duin (2004), is an unsupervised ML technique used to handle one class classification (OCC) problems. The objective of OCC is to find the boundary of the target class. In this research, the choices of buying the term deposit are considered as the target class.

Three popular methods can be applied to solve the one-class classification problems: the density estimation, the boundary method and the reconstruction methods (Tax & Duin, 2001). The basic idea of SVDD is to define a hypersphere, characterized by the center **a** and the radius R, containing as many as possible of the training samples and minimize the volume of the sphere by minimizing R^2. Figure 2 describes the difference between one class classifier and binary classifier. It is a typical optimization problem whose objective function can be expressed as:

$$\begin{gathered}\text{minimize:}\quad F(R,\mathbf{a}) = R^2 \\ \text{subject to: } \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 \le R^2\end{gathered} \tag{4}$$

where $\boldsymbol{x}_i,\ i = 1,2,\ldots,n$ denote the training dataset with n samples. In order to allow the possibility of the outliers in the training dataset, the distance from $\boldsymbol{x}_i$ should not be strictly smaller than R^2. Therefore, the slack variables are introduced to solve the problem and Equation (4) can be rewritten as:

$$\begin{gathered}\text{minimize:}\quad F(R,\mathbf{a},\xi) = R^2 + C\sum_{i=1}^{n}\xi_i \\ \text{subject to: } \|\boldsymbol{x}_i - \boldsymbol{a}\|^2 \le R^2 + \xi_i,\ \xi_i \ge 0\end{gathered} \tag{5}$$

where $\xi_i, i = 1,2,...,n$ denote the slack variables. The user-defined parameter C determines the trade-off between the volume of the sphere and the errors. More detailed descriptions are available in the research by Tax & Duin (2004).

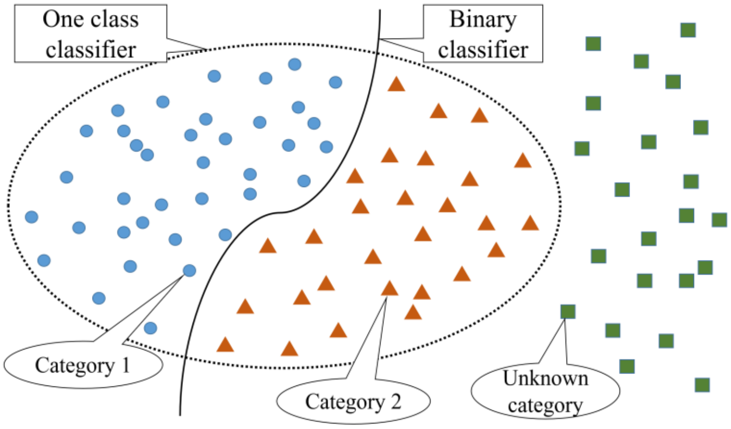
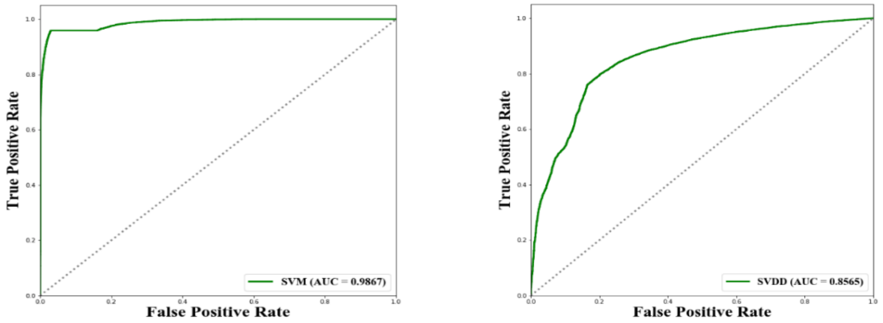


Figure 2. One class classifier and binary classifier.

3. Results and Discussions

The available dataset with regards to the bank clients' information is from UCI Machine Learning Repository contributed by Moro et al. (2014). It was collected from a Portuguese banking institution from May 2008 to November 2010. In order to save space, the detailed information of the data is unavailable. The interested readers can refer to our previous study (Gao et al., 2019).

All the experiments in this research are conducted using Python 3.60 in a Windows 10 server with an Intel Core i5 2.40 GHz processor. For the SVM model, radial basis function (RBF) is the kernel function. The penalty parameter is set to 20 and the gamma is set to 3.0 based on 'trial and error'. The dataset is split into 80% and 20% for training and validating, respectively. The training, validation as well as the overall accuracy are 97.30%, 97.279% and 97.30%, respectively. For the SVDD model, the overall accuracy is 57.30% which is just a litter higher than random guessing. The ROC curves of the data are shown in Figure 3. The AUC values for SVM and SVDD are 0.9867 and 0.8565, respectively.



(a) (b)

Figure 3. The ROC curves for (a) SVM and (b) SVDD.

4. Conclusions

A comparative study was conducted to compare the prediction performance of the SVM and SVDD. SVM outperforms SVDD based on the overall accuracy and AUC values. The results indicate that supervised learning may be more suitable for the labelled binary dataset to some degree. For the unsupervised learning, the label of the dataset is not considered. Therefore, the model may classify the samples into more than two categories which would reduce the prediction accuracy.

References

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. Journal of Retailing and Consumer Services 8(3): 147-156.

Elsalamony, H. A. (2014). Bank direct marketing analysis of data mining techniques. International Journal of Computer Applications, 85(7), 12-22.

Huang, C. L., Chen, M. C., & Wang, C. J. (2007). Credit scoring with a data mining approach based on support vector machines. Expert systems with applications 33(4): 847-856.

Hurst, S. A. (2008). Vulnerability in research and health care; describing the elephant in the room? Bioethics 22(4): 191-202.

Gao, H., Pan, X. W., Fam, P. S., & Low, H. C. (2019) Neural networks with different activation functions applied in bank telemarketing.

Mazhar, M. I., Kara, S., & Kaebernick, H. (2007). Remaining life estimation of used components in consumer products: Life cycle data analysis by Weibull and artificial neural networks. Journal of operations management 25(6): 1184-1193.

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. Decision Support Systems 62: 22-31.

Moro, S., Cortez, P., & Rita, P. (2015). Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns. Neural Computing and Applications, 26(1), 131-139.

Noble, W. S. (2006). What is a support vector machine? Nature biotechnology, 24(12), 1565-1567.

Tax, D. M., & Duin, R. P. (2001). Uniform object generation for optimizing one-class classifiers. Journal of machine learning research, 2(Dec), 155-173.

Tax, D. M., & Duin, R. P. (2004). Support vector data description. Machine learning, 54(1), 45-66.

Vapnik, V. (1998). The support vector method of function estimation. In Nonlinear modeling (pp. 55-85). Springer, Boston, MA.

Wang, L. (Ed.). (2005). Support vector machines: theory and applications (Vol. 177). Springer Science & Business Media.

Yao, X., Tham, L. G., & Dai, F. C. (2008). Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. Geomorphology, 101(4), 572-582.