Neural Network

SVM with Application to Cell Classification

Name: Sun Yuanguang

Matriculation No.: HT083121X

Email: g0803121@nus.edu.sg

Date:29/3/2009

SVM with Application to Cell Classification

Sun Yuanguang

Department of Electrical and Computer Engineering

NATIONAL UNIVERSITY of SINGAPORE

g0803121@nus.edu.sg

***Abstract*—Support vector machine (SVM) has met with significant success in numerous real-world learning tasks. In this project, SVM is implemented to classify cells into two classes: benign or malignant. Different kinds of SVMs with different kinds of margins (hard margin and soft margin) and kernels (linear kernel and polynomial kernel) are carried out. Then the classifying performances of these SVMs are compared.**

***Index Terms*—SVM, hard margin, soft margin, linear kernel, polynomial kernel**

I. INTRODUCTION

Support vector machine has strong theoretical foundations and excellent empirical successes. It has been applied to tasks such as handwritten digit recognition, object recognition, and text classification. Besides, it has been shown good performances in multiple areas of biological analysis such as evaluating microarray expression data, detecting remote protein homologies[2].

Here SVM is implemented to classify the cancer cells into two groups: benign cells and malignant cells. The data used in this project is the Wisconsin Diagnostic Breast Cancer (WDBC) Data Set. The data set consists of training data and test data. The training data includes 285 samples, each consisting of 30 real-valued features that have been computed from a digitized image of a fine needle aspirate of a breast mass. The test data consists of 100 samples, with the same structure as the samples in the training data. The training data is used to train the SVM to perform proper classification. Then the classifying performance of SVM is tested with the test data.

In this project, SVMs with different structures are carried out, including: hard-margin SVM with the linear kernel, hard-margin SVM with a polynomial kernel, soft-margin SVM with a polynomial kernel. The constraint parameters p and C are varied to investigate how they affect the performance of SVM. Training accuracy and test accuracy are used to evaluate the performance. The training accuracy is defined by the number of samples correctly classified in the training data set. And the test accuracy is defined by the number of samples correctly classified in the test data set.

The remainder of the report is structured as follows. Section 2 illustrates the principles and structures of SVM. Section 3 presents the simulation of SVM method demonstrated in section 2. Conclusions and some discussions of the results are presented in section 4.

II. Support Vector Machine

SVM is a relatively new type of learning algorithm, firstly developed by Vapnik and co-workers[2]. When used for classification, it separates the data labeled by -1 and 1 with a hyper-plane, as is shown in Figure 2.1.

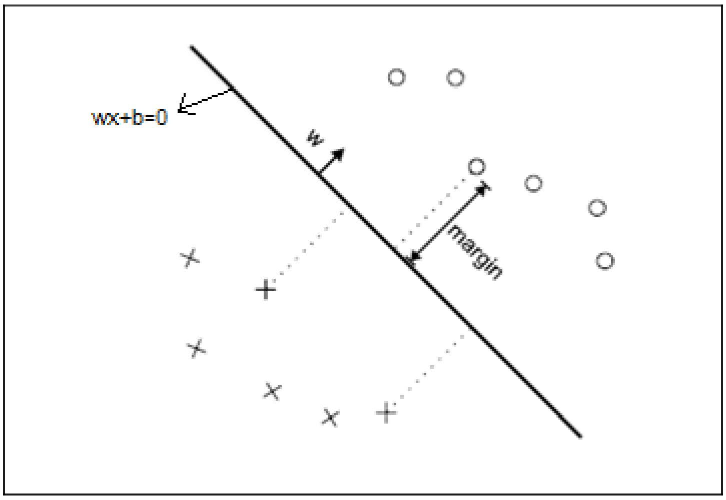

Figure 2.1 A simple linear SVM

2.1 Hard Margin SVM

The geometric margin of one sample is defined as $r_i^g = \frac{r_i^f}{\|w\|}$, where r_i^f is the functional margin of one sample. The geometric margin of the data set is defined as $r^g = \frac{r^f}{\|w\|}$, where r^f is the functional margin of the data set. The key point of SVM is to get the maximal r^g.

We can maximize r^g by fixing r^f then minimizing $\|w\|$. Thus we get the constrained optimization problem: minimizing $f(w) = \frac{1}{2}\|w\|^2$ with the condition $d_i(w^T x_i + b) \geq 1$, where d_i is the label and x_i is the feature of the data. The corresponding Lagrangian function is:

$$L(w,b,\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i (d_i (w^T x_i + b) - 1) \qquad (2.1)$$

According to Kuhn-Tuck theory, we have the KKT conditions:

$$\frac{\partial L(w,b,\alpha)}{\partial w} = 0; \quad \frac{\partial L(w,b,\alpha)}{\partial b} = 0;$$

$$d_i(w^T x_i + b) \geq 1; \quad \alpha_i(d_i(w^T x_i + b) - 1) = 0; \quad \alpha_i \geq 0$$

Extracting the above equations, we can obtain: $w_0 = \sum_{i=1}^{N} \alpha_i d_i x_i$. (2.2)

Furthermore, the original optimization problem can be transformed to the dual problem: maximizing $Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j$ with the constraints $\sum_{i=1}^{N} \alpha_i d_i = 0$ and $\alpha_i \geq 0$, for i=1,2…N.

Solving the dual problem we can obtain Lagrangian multiplier α_i. For a support vector x_s, we can obtain:

$$d_s(w_0{}^T x_s + b_o) = 1 \Rightarrow b_o = \frac{1}{d_s} - w_0{}^T x_s \tag{2.3}$$

Thus, the discrimination function is:

$$g(x) = w_0{}^T x + b_o \tag{2.4}$$

Therefore, for the input x, the output of the SVM is: $\operatorname{sgn}[g(x)]$ (2.5)

2.2 Soft Margin SVM

Compared with hard margin SVM, another constraint parameter C is introduced to soft margin SVM. The consequence procedure is similar to that used in 2.1. The primal problem is minimizing $f(w) = \frac{1}{2} w^T w + C\sum_{i=1}^{N} \xi_i$ with the condition $d_i(w^T x_i + b) - 1 + \xi_i \ge 0$ and $\xi_i \ge 0$. The dual problem is maximizing $Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i{}^T x_j$ with the constraints $\sum_{i=1}^{N} \alpha_i d_i = 0$ and $0 \le \alpha_i \le C$, for i=1,2…N.

It can be seen that the only difference from hard margin is the constraint α_i in the dual problems. In the soft margin, α_i is constrained by an upper bound C. Then we can get the same discrimination function as 2.4 and the same output function as 2.5. But we should notice that when determining b_o, we should take any data point (x_i, d_i) in the training set for which we have $0 < \alpha_i < C$. In our simulations, we take the value of b_o resulting from such data points in the training set.

2.3 Soft Margin SVM with Kernel

SVM with kernel can work in case that linear separation is impossible. The dual problem is $Q(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_i d_j K(x_i,x_j)$ with the constraints $\sum_{i=1}^{N}\alpha_i d_i=0$ and $0\le\alpha_i\le C$, for i=1,2…N. The discrimination function is:

$$g(x)=\sum_{i=1}^{N}\alpha_i d_i K(x_i,x)+b_0 \tag{2.6}$$

Note that b_o should be determined in the same way as that in 2.2.

Thus, the output of SVM is: $\mathrm{sgn}[\sum_{i=1}^{N}\alpha_i d_i K(x_i,x)+b_0]$ (2.7)

III. Simulations and Results

3.1 Simulations

In this section, three different SVMs are simulated to perform the classification task. The first one is hard-margin SVM with the linear kernel: $K(x_1,x_2)=x_1^T x_2$. The linear kernel equals to the model presented in 2.1.

The second one is hard-margin SVM with the polynomial kernel: $K(x_1,x_2)=(x_1^T x_2+1)^p$. In the simulation, the parameter p varies to be 2, 3, 4 and 5.

The third one is soft-margin SVM with the polynomial kernel. In this simulation, the parameter C varies to be 0.1, 0.6, 1.1 and 2.1 and p varies to be 2, 3, 4 and 5.

The SVM is trained by the training data and the discrimination function can be obtained. Then the training data and test data were substituted into SVM and we can get the corresponding outputs of SVM. The training accuracy and test accuracy are considered to evaluate the performance of SVM. The training accuracy is defined by the number of samples correctly classified in the training data set. And the test accuracy is defined by the number of samples correctly classified in the test data set. The results of SVM classification are shown in Table 3.1.

Table 3.1 Results of SVM classification

Type of SVM	**Training Accuracy (per 285)**				**Test Accuracy (per 100)**			
Hard Linear	285				95			
Hard Polynomial	P=2	P=3	P=4	P=5	P=2	P=3	P=4	P=5
	285	285	285	285	95	96	96	95
Soft Polynomial	C=0.1	C=0.6	C=1.1	C=2.1	C=0.1	C=0.6	C=1.1	C=2.1
P=2	281	282	285	285	97	97	96	95
P=3	285	285	285	285	96	96	96	96
P=4	285	285	285	285	96	96	96	96
P=5	285	285	285	285	95	95	95	95

3.2 Comments on the Results

It can be seen from Table 3.1 that for hard margin SVM with linear kernel, the training accuracy is 100% and the test accuracy is 95%. The total accuracy is 380/385.

When polynomial kernel is introduced, the training accuracy doesn’t change, while the test accuracy changes. When the kernel parameter p=2, the test accuracy is 95/100, which is the same as that in hard margin SVM with linear kernel. When p=3 or 4, the test accuracy is 96/100, which is a little larger. However, if p=5, the test accuracy decreases to 95/100. Furthermore, if p is chosen too large, the training accuracy and test accuracy will fall in an obvious degree. Thus we can conclude that polynomial kernel can improve the performance of SVM in case that the parameter p is not too small or too large. The best case is with training accuracy of 285/285, test accuracy of 96/100 and the total accuracy of 381/385.

In the case of soft margin SVM with polynomial kernel. When p=2, it can be seen that smaller C can improve the test accuracy. In contrary, smaller C results in lower training accuracy. Besides, if C is too small (smaller than 0.05), the training accuracy will decrease quickly in a great degree and the training accuracy will fall accordingly.

Thus we can conclude that a large C results in larger training accuracy within some scope and a small C results in smaller training accuracy.

Besides, it is shown that when p=3 and C=0.6, the training accuracy is 282/285, test accuracy of 97/100. When p=3 and C=1.1, the training accuracy is 285/285, test accuracy of 96/100. Thus is can be concluded that higher training accuracy may not lead to higher test accuracy.

Furthermore, we can find that when p is larger than 2, the training accuracy is 100% even C is small. This shows that kernel can carry out classification in the case that linear separation is impossible. Also when p is larger than 2, C can't affect the test accuracy obviously. The best case is with training accuracy of 285/285, test accuracy of 96/100 and the total accuracy of 381/385.

IV. Conclusions

In this project, SVM is implemented to classify cancer cells. The achievable highest total accuracy is 381/385. The results indicate that SVM is able to realize the classification.

It is also concluded that polynomial kernel and soft margin can improve the classifying performance of SVM. A large soft-margin parameter C leads to larger training accuracy within some scope and a small C leads to smaller training accuracy. While if C is too small (smaller than 0.05), the training accuracy will decrease quickly in a great degree. Furthermore, it is showed that kernel can carry out classification in the case that linear separation is impossible. If p is chosen properly (p=3 or 4), the training accuracy will be 100% and the test accuracy will be higher. While if p is too large, the training accuracy and test accuracy will decrease greatly.

References

1. Peter C. Y. Chen, "Lecture Note – EE5904/ME5404 Neural Networks", 2009.

2. Terrence S. Furey, Nello Cristianini, Nigel Duffy, David W. Bednarski, Michel Schummer and David Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data", Bioinformatics, Page 906-914, Vol. 16no. 10, 2000.