### The mathematical background behind the SVM

SVM is a two-dimensional description of the optimal surface evolved from the linearly separable case, the basic idea can be used in figure 1. Figure 1 shows the basic idea. Two types are separated by H without errors. H1 and H2 are places which pass the recent point of H. The distance of H1 and H2 was called class interval. Optimal separating surface is not only to ensure error-free separation of the two types of samples, also called the largest class interval.

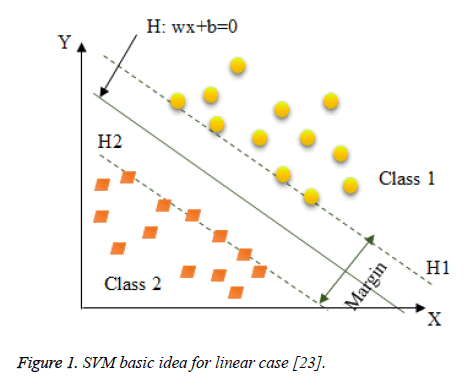


Figure : Optimal separating surface

In pattern recognition, the linear discriminate function in the n-dimensional space:

The classification hyperplane equation can be written . Linear separable case, the discriminate function was normalized. So that all training samples are met , even to meet away from the classification of the surface of the sample , so the class interval equivalent to , thus making the interval on the equivalent to or . Make a classification of the surface of all samples correctly classified, it is necessary to meet:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Satisfy the above equation (1) and make the smallest classification surface is the optimal classification surface. A point on the hyperplane are called support vectors, they support the optimal classification surface (Deng N., 2009).

#### Linearly Separable

According to the above, the optimal separating surface can be expressed as the following constrained optimization problem, Seeking the minimum of the following the function:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

We can define the Lagrange function as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Where is a coefficient of Lagrange. Problem is seeking the minimum of and to the Lagrange fn. and seek partial differential and make them equal to zero, the original problem can be transformed into the following simple dual problem:

the constraints

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

|  |  |  |
| --- | --- | --- |
|  | ,n | (5) |

The maximum value of solving the following function about :

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

If is the Optimal solution,

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

The optimal weight coefficient vector of the surface is a linear combination of the training sample vector. According to the Kuhn-Tucker conditions, the optimal solution must also meet the following conditions:

|  |  |  |
| --- | --- | --- |
|  |  |  |

For most samples will be zero, the value of zero correspondings to the type with equality (1) are support vector, they are usually all the samples part. To solve the above problems, the optimal Category function is

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

sgn() is the sign function.

#### Linear Inseparable Problem

The main idea of the linearly inseparable problem, the support vector machine is the input vector is mapped to a high-dimensional feature vector space and constructs the optimal separating surface in the feature space. When the training set of linear nontime-sharing, some training samples can not satisfy condition (1). Conditions can be amended to add a slack variable .

The constraints are:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

When misclassification occurs, , is training set the upper bound of the number of misclassified. It can be used as a description of the degree of the training set is misclassified.

This makes it two goals: is large as possible. is as small as possible.

These two goals into one goal, the introduction of the penalty parameter C as the right combination of these two target weight, under the constraints of the conditions (9) and the minimum of the following functions:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

Where: C is the penalty parameter. It actually plays a role in the control of the degree of right or wrong sub-sample of punishment to achieve a compromise between the proportion of misclassified samples and the complexity of the algorithm. Its optimal solution to the saddle point of Lagrange function is defined as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

|  |  |  |
| --- | --- | --- |
|  |  |  |

Where and is all Lagrange multiplier vector. The specific method for solving Linear separable is very similar, but the condition is changed (N., 2011; Zhang Y., 2011; Zhang, 2012)

| **meka danna methods walata**

The experiments show two explanations of why the method for grid search has been chosen

1. The likelihood of feeling unsafe by using methods to avoid a thorough search of approximation parameters.
2. In order to find good Grid Search parameters, the computational time required by advanced methods is not much greater because only two parameters are usable.