**Support Vector Machine-Duality Problem**

[[](https://medium.com/@kamil2000budaqov?source=post_page-----10442bb2f6cc--------------------------------)](https://medium.com/@kamil2000budaqov?source=post_page-----10442bb2f6cc--------------------------------)

[[Becoming Human: Artificial Intelligence Magazine](https://becominghuman.ai/?source=post_page-----10442bb2f6cc--------------------------------)](https://becominghuman.ai/?source=post_page-----10442bb2f6cc--------------------------------)

[Kamil Budagov](https://medium.com/@kamil2000budaqov?source=post_page-----10442bb2f6cc--------------------------------)

·

[Follow](https://medium.com/m/signin?actionUrl=https%3A%2F%2Fmedium.com%2F_%2Fsubscribe%2Fuser%2F74baf91e26f&operation=register&redirect=https%3A%2F%2Fbecominghuman.ai%2Fsupport-vector-machine-duality-problem-10442bb2f6cc&user=Kamil+Budagov&userId=74baf91e26f&source=post_page-74baf91e26f----10442bb2f6cc---------------------post_header-----------)

Published in

[Becoming Human: Artificial Intelligence Magazine](https://becominghuman.ai/?source=post_page-----10442bb2f6cc--------------------------------)

·

9 min read

·

Oct 21, 2022

25

1

**Introduction**

In this paper, I will discuss the Support Vector Machine (SVM), which is a supervised learning algorithm for classifying data. Generally, SVM is used for separating multidimensional data points into two different classes. For simplicity, we will consider only the two-dimensional case (the idea for higher dimensionality is the same). First, we will define the objective function subject to constraints and then consider the Lagrangian dual function for solving this optimization problem. In addition, I will prove that solving the dual problem is equivalent to solving the primal problem.

**Support Vector Machine**

Suppose that we have the following data set, and we are required to find the best line that separates the data into two classes:

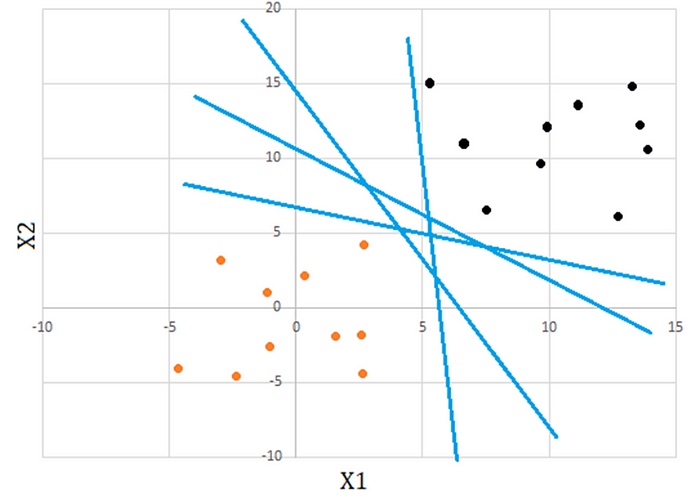


Image by author. Figure 1

As it is indicated in Figure 1, there are many lines that can do this job. However, our criteria for the best line will be as follows:

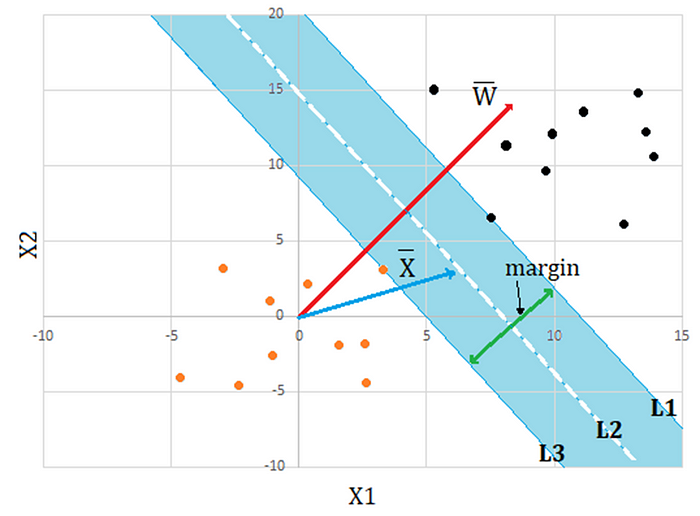


Image by author. Figure 2

The distance from line*L1* to *L2* is equal to that from *L2* to *L3*. The vector *W* is perpendicular to all lines, namely *L1*, *L2*, and *L3*. The objective here is to maximize the margin between lines *L1* and *L3*. Additionally, it should be noted that lines *L1* and *L3* must contain at least one point from the appropriate class. After obtaining the equation for the line *L2*, it is easy to classify new data points.

As it is clear from the picture, there is a functional dependency between the vector *W* and the margin. Systematically, we will construct this dependency and then try to solve the problem.

Several conclusions can be drawn from Figure 2. First, the dot product of vector *W* with any vector *X* lying on the line *L2* is independent of the position of the point on the line and can be denoted by some constant *C*



Assuming that constant *B =-C*, we can rewrite the above formula as:



Equation 1.1

This is actually an equation of *L2*.

We can think of Equation 1.1 as a level curve of



when the function is set equal to zero

Apparently, the vector *W* is the gradient of *F (X),* implying that the function is increasing in the direction of W**.**Now we can conclude that for points above *L2*, the value of the function will be positive and for those below it will be negative. We want *F (X)* to be such a function for which the level curves at *F (X) = 1*and*F (X) =-1*produce lines L1 and L3, respectively.

Similar to *L2*, for lines *L1*and *L3*, the dot product of *W* with any vector *X* lying on the respective lines can be shown as:



Equation of L1 and L3

By exploiting the fact that the function is increasing along the gradient, we can say that, for points above *L1*, the following inequality holds:



Inequality 1.1

And for those below line *L3*:



Inequality 1.2

As indicated in Figure 1, we actually know to which class our given points belong. (indicated by black and orange colors). Let us define a scalar variable *yi* for the 2D point *xi* . For the points that satisfy inequality 1.1, let the variable *yi* be +1, indicating one class, and for those that satisfy inequality 1.2 , *yi* is -1, indicating another class.



In both cases, the following inequality holds:



Inequality 1.3

From now on, inequality 1.3, which holds for all points, will be our constraint.

Now let us consider the following picture in order to derive the formula for our margin:

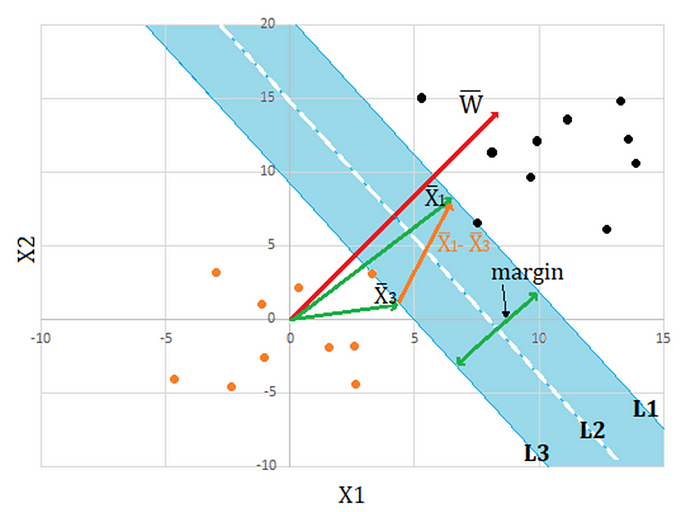


Image by author. Figure 3

The margin can be calculated as a dot product of *X1-X3* with a unit vector in the direction of *W.*



From the equations for*L1* and *L3*:



Therefore, the margin will be



Eventually we can state problem like maximizing this margin subject to constraints in inequality 1.3

For mathematical convenience, we will minimize



subject to inequality 1.3, which can be rewritten as:



For now, let us leave this problem here and introduce the Lagrangian dual function for solving an optimization problem with inequality constraints.

**Lagrangian Duality**

In order to understand the idea of duality, let us first generalize the problem by restating it like this:

Minimize



Subject to the following inequality constraints:



Where *x* is a point in multidimensional space

This is our primary problem. Let me denote this problem by *P1*. From now on, I will refer to it as *P1*. Now let us finally introduce the Lagrangian dual function, which transforms *P1* as follows:



The dual problem is written as follows:



From now on, we will refer to this problem with the name “dual problem." First, we will show that the solution to the dual problem is less than or equal to the solution to the primal problem. This is called **“weak duality"** and can be expressed mathematically as in the following:



**Proof of Weak Duality**

In *P1*, if for some *x*, at least one of the constraints is violated, then the supremum of *L* with respect to lambda will be positive infinity. In mathematical form, it can be shown as follows:

In *P1*, if x violates at least one of the constraints



Otherwise (if none of the constraints is violated):



We can draw following conclusions, assuming that ***P\**** is a solution to *P1*, and ***x\**** is a point where ***F(x\*) = P\****



Therefore, we can rewrite weak duality as:



Inequality for weak duality

(Note that the maximum or minimum of a function is a finite number, while supremum and infimum can be infinite).

In order to prove the inequality for weak duality, let us first show that following inequality holds



In extended form:



Just by substituting *x\** to the left, we can prove that:



This inequality holds for x\* because of the following conditions:



Since inequality holds for *x\*,*the infimum will definitely be less than *P\*.*

When evaluating the infimum of *L* for *x* , we assume that lambda is fixed and *x* is varying. Therefore, for all nonnegative lambdas, weak duality is satisfied, including the values of lambda which maximize this infimum. This can be formulated as follows:



**Strong Duality**

In a special case, when the objective function and all constraints are convex(as in our SVM problem), strong duality may hold, which says that:



We can rewrite strong duality as follows by denoting the infimum of *L* for *x*by the function of lambda:



Now let us denote the solution to the left side by *D\*,* which function *q* evaluates at lambda star.



We also know that



because, from the definition of infimum, it is the minimum possible value and can be at most equal to the right hand side of the above inequality. Considering this fact,



Just by simplifying the above inequality, we end up with:



And it is only possible if:



I would like to emphasize that each term of summation can be a negative number or zero. Therefore, the necessary condition ( called complementary slackness) for the entire sum to be zero is that, in fact, all terms must be zero, as shown below:



We can deduce that lambda values associated with negative valued constraints must be zero. In terms of SVM, lambda values associated with points which do not lie on lines *L1* and *L3* must be zero. Intuitively, the addition or removal of these points should not change the margin.

**SVM problem**

Now let us come back to our original problem and try to solve it by introducing a dual problem.

Minimize



Subject to following constraints



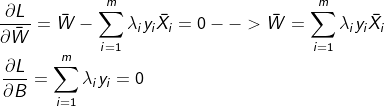
The Lagrangian extension will be



And dual problem:



We must start by finding the infimum, which is done by finding partial derivatives with respect to W and B and setting them equal to zero:



Substituting these equations back to the dual problem, the dual problem will be:



with following constraints:



**Conclusion**

In this paper, I explained how the constrained optimization problem can be transformed into the dual problem with the introduction of Lagrangian extension. We also derived a necessary condition called **“complementary slackness,”**assuming that strong duality holds**.**Finally, we can say that if we have an optimal solution for a dual problem, we also have an optimal solution for a primary problem, or vice versa. The optimal solution to one of the problems proves that our assumption is correct (strong duality holds). In other words, the duality gap between primal and dual problems is zero. I hope the SVM example was a good example to demonstrate how the dual problem can be applied to solve a real classification problem. In the end, I would like to emphasize that the same approach can be applied to find the equation of hyperplane to separate multidimensional data points into two classes.