SVM

14 March 2023

Sumit Kumar Yadav

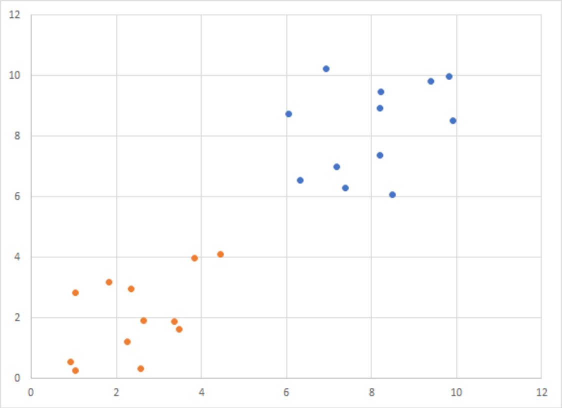
Department of Management Studies
Indian Institute of Technology, Roorkee

Recap and Today

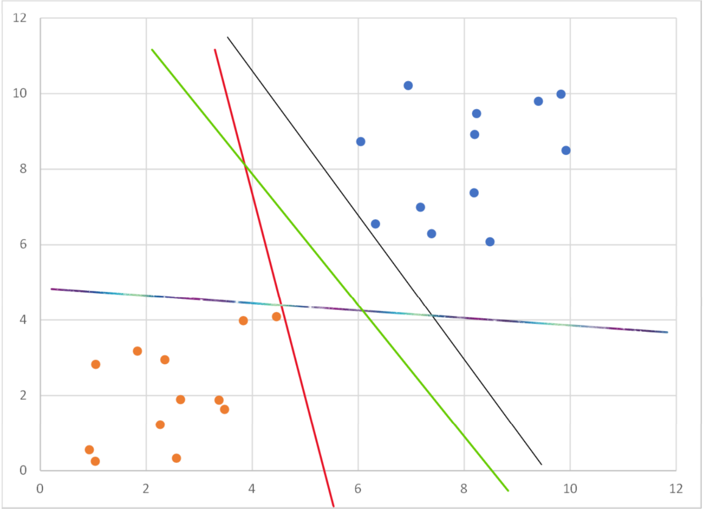
- SVM
- More of SVM

Support Vector Machine - Introduction

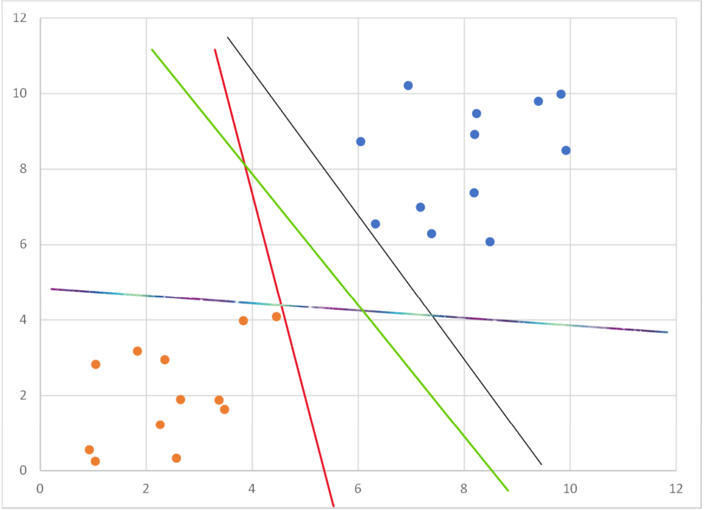
- Supervised Learning Algorithm for Classification
- Given N training points in which n_1 are of type A, n_2 are of type B, draw the **best** line(plane)
- To begin with, assume that the training points are linearly separable



Which is the best line?



Which is the best line?



What makes us think it is the green line? Can we make the ideas a bit more precise?

Notations

Let the data-set be denoted as -

S.No	X_1	X_2	Y (+1 or -1)
1	x_{11}	x_{12}	+1
2	x_{21}	x_{22}	-1
3	x_{31}	x_{32}	-1
. . .	. . .	. . .	. . .
N	x_{N1}	x_{N2}	+1

Notations

Let the data-set be denoted as -

S.No	X_1	X_2	Y (+1 or -1)
1	x_{11}	x_{12}	+1
2	x_{21}	x_{22}	-1
3	x_{31}	x_{32}	-1
. . .	. . .	. . .	. . .
N	x_{N1}	x_{N2}	+1

Let the equation of the line be $w_1 x_1 + w_2 x_2 + b = 0$
We need to determine w_1, w_2 and b

Obtaining w_1, w_2 and b

- Arbitratily choose w_1, w_2 and b such that
  $w_1 x_{i1} + w_2 x_{i2} + b > 0$ whenever $y_i = +1$
  $w_1 x_{i1} + w_2 x_{i2} + b < 0$ whenever $y_i = -1$
- Simply put, for all points, $y_i(w_1 x_{i1} + w_2 x_{i2} + b) > 0$
- Criteria - Consider a line. Find the distance of the line from all the training examples (or points). Look at the minimum of all these distances.
  We are interested in the line for which this minimum distance is as large as possible.
- $$\max_{(w_1, w_2, b)} \left(\min_{i=\{1,2,\ldots,N\}} \frac{|w_1 x_{i1} + w_2 x_{i2} + b|}{\sqrt{w_1^2 + w_2^2}} \right)$$
- The **max** in the equation in **maximize** and **min** in the equation is **minimum**

Optimization Problem

The optimization problem thus becomes -

$$\max_{(w_1,w_2,b)} \left(\min_{i=\{1,2,....,N\}} \frac{|w_1 x_{i1} + w_2 x_{i2} + b|}{\sqrt{w_1^2 + w_2^2}} \right)$$

The **max** in the equation in **maximize** and **min** in the equation is **minimum**

subject to the following **N** constraints - $y_i(w_1 x_{i1} + w_2 x_{i2} + b) > 0$

Optimization Problem

The optimization problem thus becomes -

$$\max_{(w_1,w_2,b)}\left(\frac{1}{\sqrt{w_1^2+w_2^2}}\left[\min_{i=\{1,2,....,N\}}(|w_1x_{i1}+w_2x_{i2}+b|)\right]\right)$$

subject to the following **N** constraints -

$$y_i(w_1x_{i1}+w_2x_{i2}+b)>0$$

Optimization Problem

The optimization problem thus becomes -

$$\max_{(w_1,w_2,b)} \left(\frac{1}{\sqrt{w_1^2 + w_2^2}} \left[\min_{i=\{1,2,....,N\}} (|w_1 x_{i1} + w_2 x_{i2} + b|) \right] \right)$$

subject to the following **N** constraints -

$$y_i(w_1 x_{i1} + w_2 x_{i2} + b) > 0$$

As scaling all w_1, w_2 and b by the same factor (non-zero) doesn't change the line (or hyperplane), we will choose w_1, w_2 and b such that -

$\min_{i=\{1,2,....,N\}}(|w_1 x_{i1} + w_2 x_{i2} + b|) = 1$

The **min** in the above equation is **minimum**

Optimization Problem

The optimization problem thus becomes -

$$\max_{(w_1,w_2,b)} \left(\frac{1}{\sqrt{w_1^2 + w_2^2}} \right)$$

subject to the following **2N** constraints -

$$y_i(w_1 x_{i1} + w_2 x_{i2} + b) > 0$$

$$\min_{i=\{1,2,....,N\}}(|w_1 x_{i1} + w_2 x_{i2} + b|) = 1$$

Optimization Problem

The optimization problem thus becomes -

$$\min_{(w_1,w_2,b)} \left(\frac{w_1^2 + w_2^2}{2} \right)$$

subject to the following **2N** constraints -

$$y_i(w_1 x_{i1} + w_2 x_{i2} + b) > 0$$

$$\min_{i=\{1,2,....,N\}}(|w_1 x_{i1} + w_2 x_{i2} + b|) = 1$$

Optimization Problem

The optimization problem thus becomes -

$$\min_{(w_1,w_2,b)} \left(\frac{w_1^2 + w_2^2}{2} \right)$$

subject to the following **2N** constraints -

$$y_i(w_1 x_{i1} + w_2 x_{i2} + b) > 0$$

$$\min_{i=\{1,2,....,N\}}(|w_1 x_{i1} + w_2 x_{i2} + b|) = 1$$

$\min_{i=\{1,2,....,N\}}(|w_1 x_{i1} + w_2 x_{i2} + b|) = 1$ implies -
$|w_1 x_{i1} + w_2 x_{i2} + b| >= 1 \qquad \forall i = \{1, 2, ..., N\}$ or

$|y_i(w_1 x_{i1} + w_2 x_{i2} + b)| >= 1 \qquad \forall i = \{1, 2, ..., N\}$ or
$y_i(w_1 x_{i1} + w_2 x_{i2} + b) >= 1 \qquad \forall i = \{1, 2, ..., N\}$ The implies condition is not both ways, but still it can be replaced in this problem because ??

Optimization Problem

The implies condition can be replaced in this problem because ??
After some algebra, the optimization problem becomes -

$$\min_{(w_1,w_2,b)} \left(\frac{w_1^2 + w_2^2}{2}\right)$$

subject to the following **N** constraints -

$$y_i(w_1 x_{i1} + w_2 x_{i2} + b) >= 1 \qquad \forall i = \{1, 2, ..., N\}$$

Why the name support vector?

Consider the following two optimizaion problems -

$$\min_{(w_1,w_2,b)} \left(\frac{w_1^2+w_2^2}{2}\right)$$

subject to the following **N** constraints -

$$y_i(w_1x_{i1}+w_2x_{i2}+b) >= 1 \qquad \forall i = \{1,2,...,N\}$$

$$\min_{(w_1,w_2,b)} \left(\frac{w_1^2+w_2^2}{2}\right) - \sum_{i=1}^{N} \alpha_i(y_i(w_1x_{i1}+w_2x_{i2}+b)-1)$$

subject to no constraints, only the fact that all α_i's are either zero or positive

Which of these two optimization problems has a lower value?

Why the name support vector?

$$\min_{(w_1,w_2,b)} \left(\frac{w_1^2 + w_2^2}{2}\right)$$

subject to the following **N** constraints -

$$y_i(w_1x_{i1} + w_2x_{i2} + b) >= 1 \qquad \forall i = \{1, 2, ..., N\}$$

$$\min_{(w_1,w_2,b)} \left(\frac{w_1^2 + w_2^2}{2}\right) - \sum_{i=1}^{N} \alpha_i(y_i(w_1x_{i1} + w_2x_{i2} + b) - 1)$$

subject to no constraints, only the fact that all α_i's are either zero or positive

Let us say that the optimization problem in blue box is optimal for $w_1 = w_1^*, w_2 = w_2^*$ and $b = b^*$. The value of the optimization problem in red box is lower at these values. Thus, the one in the red box may have a further lower optimal value. Now, let us keep playing with putting different values of α_i's.

Why the name support vector?

Let us keep playing with putting different values of α_i's and try to solve the following optimization problem.

$$\max_{\alpha_i}\left[\min_{(w_1,w_2,b)}\left(\frac{w_1^2+w_2^2}{2}\right)-\sum_{i=1}^{N}\alpha_i(y_i(w_1x_{i1}+w_2x_{i2}+b)-1)\right]$$

subject to the constraint that all α_i's are either zero or positive

The following can be shown, with some difficulty (we will not be looking at the proof of this). Refer KKT conditions.

- The value of the optimization problem above will be the same as the value of the optimization problem in the blue box in the previous slide
- The optimal value will be attained for $w_1 = w_1^*, w_2 = w_2^*$ and $b = b^*$

Why the name support vector?

$$\max_{\alpha_i}\left[\min_{(w_1,w_2,b)}\left(\frac{w_1^2+w_2^2}{2}\right)-\sum_{i=1}^{N}\alpha_i(y_i(w_1x_{i1}+w_2x_{i2}+b)-1)\right]$$

subject to the constraint that all α_i's are either zero or positive

The inner optimization problem can be solved like a usual minimization problem with no constraints. We take partial derivative with respect to w_1, w_2 and b to get the following:

- $w_1 = \sum_{i=1}^{n} \alpha_i y_i x_{i1}$ and $w_2 = \sum_{i=1}^{n} \alpha_i y_i x_{i2}$
- $\sum_{i=1}^{n} \alpha_i y_i = 0$

Only the training points for which α_i is non-zero contribute in deciding the value of w_1 and w_2. These points are called support vectors.

The new optimization problem

On substituting the conditions, w_1, w_2 and b disappear from the inner optimization problem.

$$\max_{\alpha_i} \left[\sum_{i=1}^{n} \alpha_i \; - \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_{i1} x_{j1} + x_{i2} x_{j2}) \right) \right]$$

subject to the following **N+1** constraints that

$$\alpha_i >= 0$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

The non-separable cases - Kernel Trick

$$\max_{\alpha_i} \left[\sum_{i=1}^{n} \alpha_i \; - \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\vec{x_i}.\vec{x_j}) \right) \right]$$

subject to the following **N+1** constraints that

$$\alpha_i >= 0$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

Thinking of the points as vectors.

Thank you for your attention