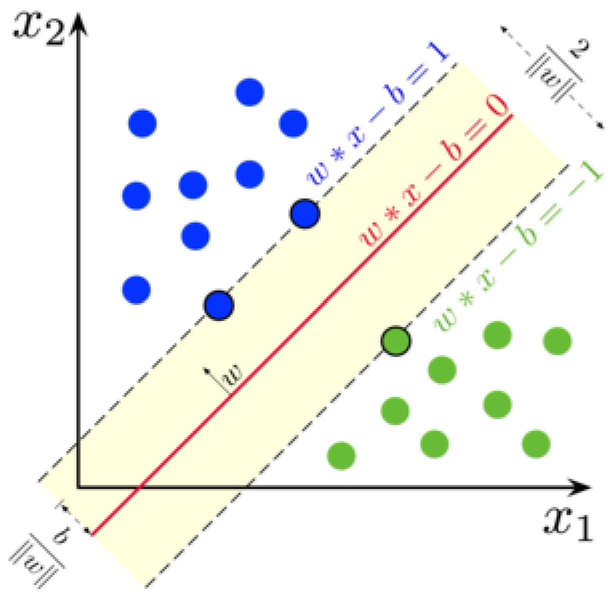
Support Vector Machines (SVM): The Coordinate Geometry Intuition

[Support Vector Machines (SVM): The Coordinate Geometry Intuition | by Ketan Suhaas Saichandran | Artificial Intelligence in Plain English (medium.com)](medium.com)

This article is for people who already have some basic understanding of SVMs and would like to understand the geometric aspects of the algorithm.



Wikipedia

So, we know the problem is about finding the best hyperplane that maximizes the margin. To make it easier to visualize, let's assume that the data vector has two dimensions. Let X be the data vector, W be the weights vector of the hyperplane and c be the constant offset. The corresponding hyperplane would be,

$$W^T X + c = 0$$

Why exactly do we choose our support vectors to lie on the following planes?

$$W^T X + c = 1, W^T X + c = -1$$

equation-1

Let’s break the vector geometry down into coordinate geometry.

$$W = [a, b]^T, X = [x, y]^T$$

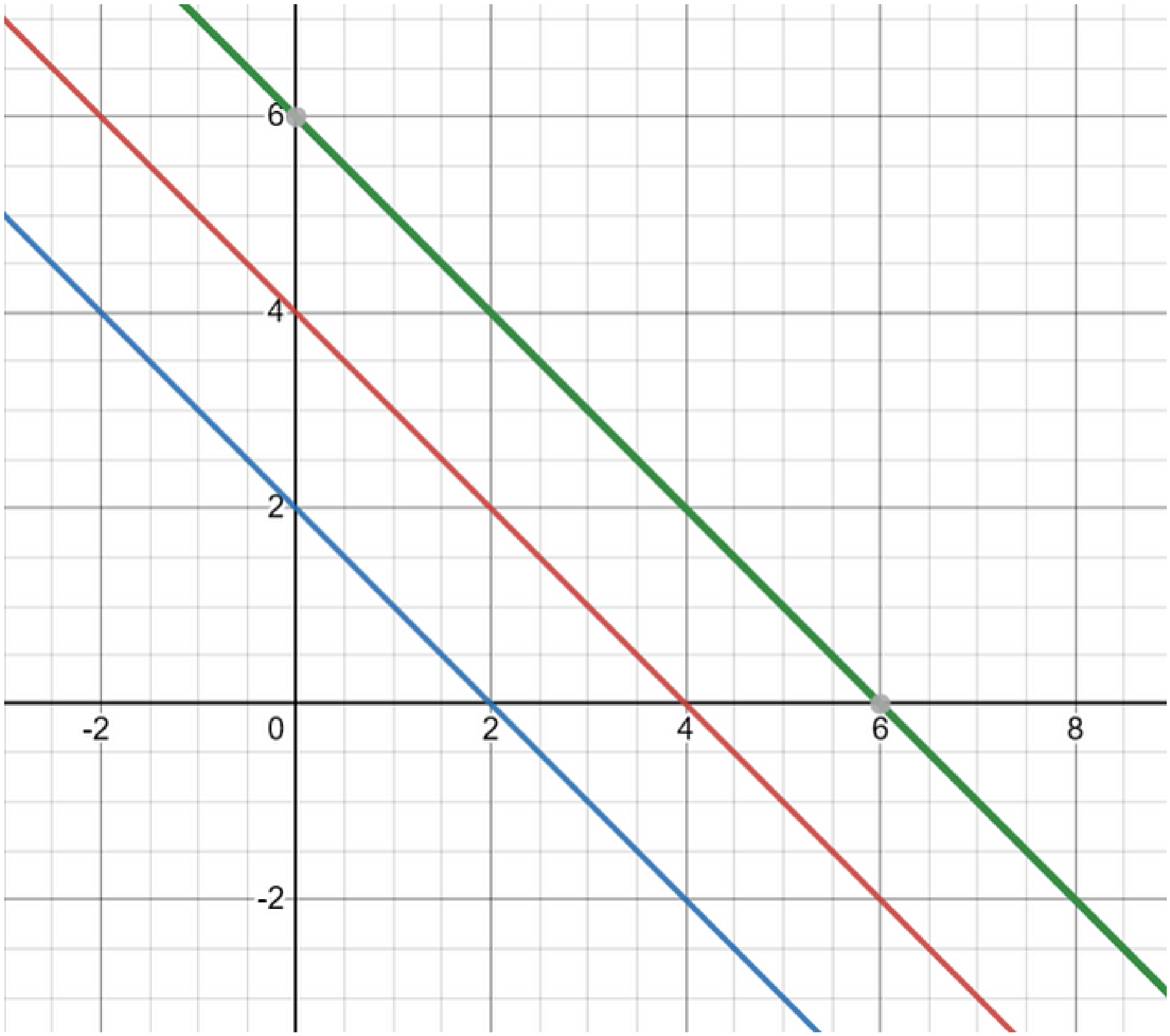
Now let’s form the hyperplane equation again.

$$ax + by + c = 0$$

Let g(x, y) be the equation of the line.

$$g(x, y) = ax + by + c$$

Analogously, g(x, y) = W^T.X + c. If you change the value of c, you are basically shifting the line closer and away from the origin.



Example

Basically, the support vectors lie on,

$$W^T.X + (c-1) = 0, W^T.X + (c+1) = 0,$$

But again, why did we choose these equations? i.e., choose 1 on RHS?

Let’s write a general form of the line equations on which the support vectors lie,

$$ax + by + c = p, ax + by + c = -p$$

Rearranging, we get ax + by + (c-p) = 0 and ax + by + (c+p) = 0. It's just shifting the line up and down, as only the y-intercept is being modified.

Now, divide both these equations by p,

$$\frac{a}{p}x + \frac{b}{p}y + \frac{c}{p} = 1, \frac{a}{p}x + \frac{b}{p}y + \frac{c}{p} = -1,$$

If you observe closely a, b and c are the learnable parameters here. And when you divide the hyperplane by p,

$$\frac{a}{p}x + \frac{b}{p}y + \frac{c}{p} = 0,$$

It's the same hyperplane equation! Now set a* = a/p , b* = b/p and c*=c/p.

$$a^*x + b^*y + c^* = -1, a^*x + b^*y + c^* = 0, a^*x + b^*y + c^* = 1$$

Convert them back to the vector form,

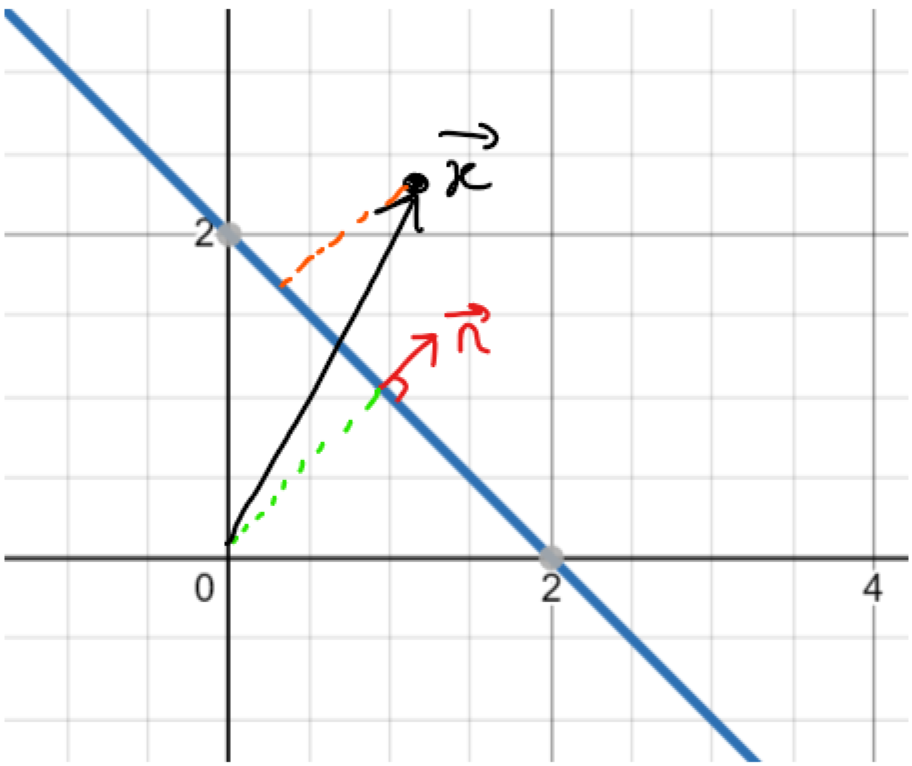
$$W^{*T}X + c = -1, W^{*T}X + c = 0, W^{*T}X + c = 1$$

Now, we have understood why equation-1 was a valid choice.

This implies that only by changing the W vector and the c offset, we can manipulate the line equation in such a way that we can fix RHS = 1. It helps us to decrease the number of parameters that we are trying to optimize.

Extra:

How do we calculate the distance of a point from a line? If x = [x1, x2]^T is a point in space, q = [q1, q2]^T is a point on the line and n is the normal vector to the line,



The orange line (d) is the distance of the point from the line.

$$x.\hat{n} - q.\hat{n} = d$$

equation-2

Or more simply,

$$(x-q).\widehat{n}=d$$

The difference of distance of x from origin in the direction of the normal vector and distance of q from origin in the direction of the normal vector.

Break equation-2 down into coordinate geometry and since q is constant,

$$\frac{xn_1+yn_2-constant}{\sqrt{n_1^2+n_2^2}}, n=[n_1,n_2]$$

That's just,

$$\frac{ax+by+c}{\sqrt{a^2+b^2}}$$

The sign of this gives the side on which the point lies on.