## Building a margin error

Recall that the margin is the distance *between the two lines*, and we want to turn this margin into an error that we can minimize using gradient descent. We want a function that gives us a small error for the large margin case, and a large error for the small margin case. This is because we want to punish small margins, as our goal is to obtain a model that has as large a margin as possible.

We have our line with the two other boundary lines, and the margin is the distance between the two outside lines.

Our equation is a line:

* *Wx*+*b*=0

and the two dotted lines have equations:

* *Wx*+*b*=1
* *Wx*+*b*=−1

The margin is 2 divided by the norm of the vector W 2 / ∣*W*∣

​

Remember that the norm of W is the square root of the sum of the squares of the components of the vector, which are W1 and W2.

For this error, let's find something that gives us

* A large value if the margin is small
* A small value if the margin is large

The norm of W (∣*W*∣) appears in the denominator. If we take the norm of W, that grows inversely proportional to the margin. To avoid dealing with square roots, let's take the norm of W squared, which is actually the sum of the squares of the components of the vector W. In this case, it's

*w*1 ^ 2+*w*2 ^ 2.And as we've seen, since W appears here in the denominator, then a large margin gives us a small error and a small margin gives us a large error. *That is exactly what we wanted*.

To clarify things, here's an example.

Let's say *W*=(3,4) and our bias is 1. So our equation of the form *w*1*x*1+*w*2*x*2+*b*=0, is going to be 3*x*1+4*x*2+1=0, and that's our main line. And the two companion lines, 3*x*1+4*x*2+1=1, and 3*x*1+4*x*2+1=−1. The error is ∣*W*∣ ^ 2, which is 3 ^ 2 + 4 ^ 2 and gives us an error of 25. The margin is 2 / ∣*W*∣, and the ∣*W*∣ is the square root of 25 which is 5. So, the margin is 2/5 and the error is 25. Let's remember these two numbers: error:25 and margin:2/5.

Now, let's look at a very similar example

Instead of our previous weights, let's assume *W*=(6,8) and our bias is 2. Our line is going to have the equation: 6*x*1+8*x*2+2=0. If you notice, that equation is the same as before except multiplied by 2. So, it gives us the same boundary line because when

3*x*1+4*x*2+1=0, then 6*x*1+8*x*2+2=0. But now our dotted lines are closer to each other. Before we had 3*x*1+4*x*2+1=1. And now, we have the twice of that equals one, which means

3*x*1+4*x*2+1 is actually 1/2, which means the line is much closer. It's actually half the distance as before, and the same thing happens with the line below.

Our error is a square of the norm of this vector, which is 6 ^ 2 + 8 ^ 2 which is 100. And our distance is going to be 2 / ∣W∣, which is 2/10. That is the same as 1/5, so this is smaller than the previous margin of 2/5. Two model examples give us the same boundary line, but one of them gives us a larger margin than the other one.

## Summary

We have our large margin, our margin of 2/5 that gives us a small error of 25, and our small margin of 2/10, which is 1/5 gives us a larger error of 100.

That is the margin error. It's just ∣*W*∣ ^ 2. This is the exact same error that is given by the regularization term in L2 regularization.

### Quiz Question

Which of the following are true about SVM ( There’s more than one correct answer )

1. Large margin = small error
2. Large margin = large error
3. Small margin = large error
4. Small margin = small error