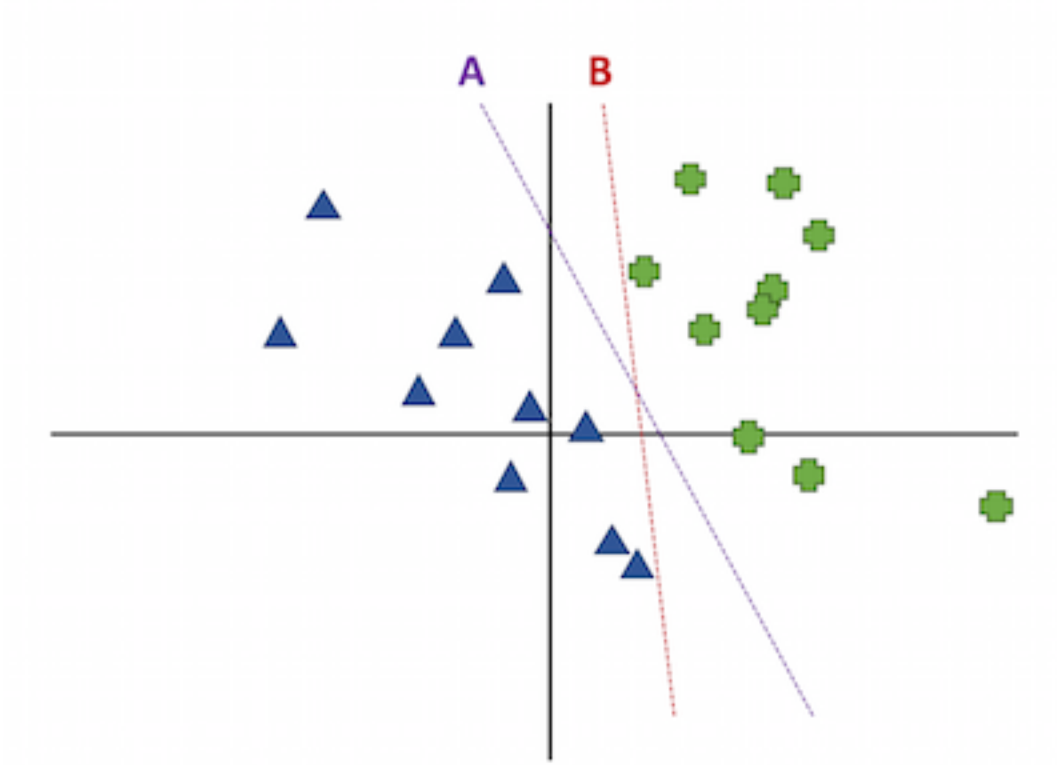
Boundaries, hyperplanes, and slopes

In the lecture on support vector machines, we looked at different decision boundaries in 2D plots like this:



In this image, line A has a larger margin than line B because it has more space separating it from the nearest points. A also has a smaller ||**w**||, because when we learned about SVMs we learned that smaller weights give larger margins.

You might be wondering how the weights **w** relate to the line that is shown. Earlier in the semester I said that the weights represent the slope of the hyperplane, and you can visually see that B has a larger slope in this plot -- so why doesn't B have larger weights?

Let’s discuss how the weights **w** relate to the slope of the decision boundary.

The lines you see on the plot above are not the hyperplane **w**T**x**. One realization to have is that a line only has one independent variable (y=mx+b), whereas in this illustration, the instances actually have two features, so x and y are both independent variables for the classifier. Instead of writing **w**T**x**, let's write out the expanded equation for the hyperplane, using both x and y as the names of the features: $w_1x + w_2y + b$. (Remember that 'b' is the intercept, which I usually leave out of the notation, but it's still there.)

This equation has two independent variables which makes it a plane, not a line. So why do you see just a line in the plot of the decision boundary? The decision boundary isn't just the plane $w_1x + w_2y + b$, but specifically the boundary $w_1x + w_2y + b = 0$. It's the "slice" of the plane where it passes through 0, which forms a line. You can also see this algebraically by rewriting this as:

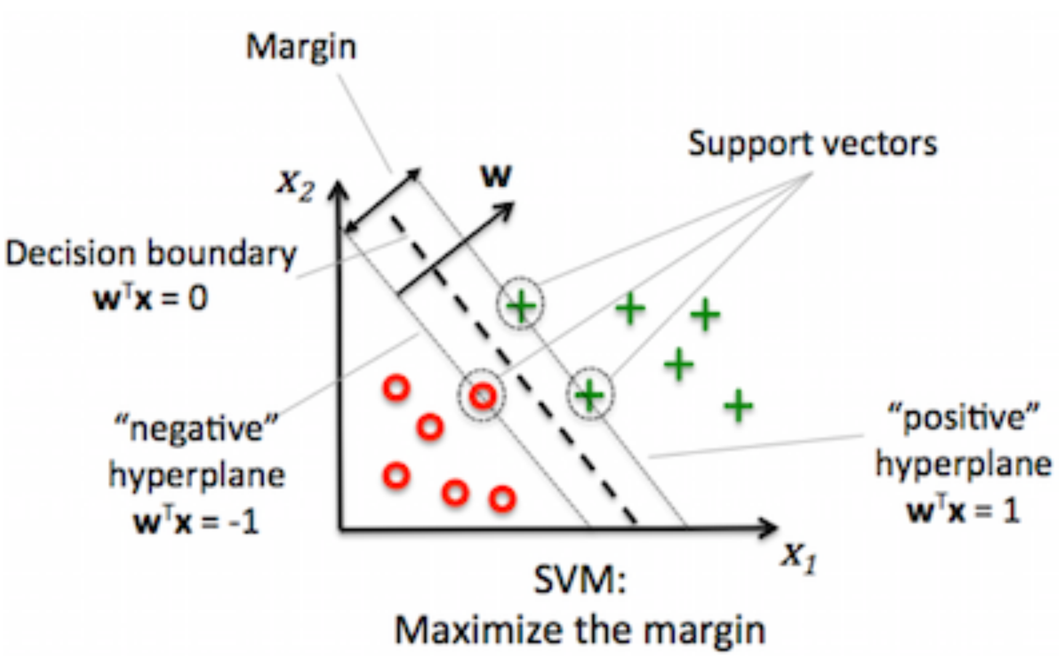
$$w_1x + w_2y + b = 0$$
$$w_2y = -w_1x - b$$
$$y = \frac{-w_1x-b}{w_2} = \frac{-w_1}{w_2}x - \frac{b}{w_2}$$

Now we have a line in the form of y=mx+b, where the slope corresponds to $\frac{-w_1}{w_2}$ and the y-intercept corresponds to $\frac{-b}{w_2}$. This line is the decision boundary that you see plotted. While the slope is based on the weights $(\frac{-w_1}{w_2})$, it's different from the slope of the full plane that defines the classifier scores, $w_1x + w_2y + b$.

This all applies to more dimensions. In general, there is a hyperplane of K dimensions that defines the score of the classifier. The decision boundary is the set of points of that hyperplane that pass through 0 (or, the points where the score is 0), which is going to be a hyperplane with K-1 dimensions.

Now let me explain why smaller weights lead to larger margins.

Remember in an SVM, instead of one decision boundary **w**T**x**=0, we have two boundaries, **w**T**x**=1 and **w**T**x**=-1, illustrated like this:



Following the same steps from earlier in this post, let's rewrite the boundaries **w**T**x**=1 and **w**T**x**=-1 in full, using x and y as the variables.

The positive boundary is:

$$w_1 x + w_2 y + b = 1$$
$$w_2 y = 1 - w_1 x - b$$
$$y = \frac{1 - w_1 x - b}{w_2} = \frac{-w_1}{w_2} x + \frac{1-b}{w_2}$$

The negative boundary is:

$$w_1 x + w_2 y + b = -1$$
$$w_2 y = -1 - w_1 x - b$$
$$y = \frac{-1 - w_1 x - b}{w_2} = \frac{-w_1}{w_2} x + \frac{-1-b}{w_2}$$

Both of these boundaries are lines with the same slope, so they are parallel. The margin is the distance between these two parallel boundaries, which turns out to be:

$$\frac{2}{\sqrt{w_1^2 + w_2^2}}$$

(Why? See https://en.wikipedia.org/wiki/Distance_between_two_straight_lines)

Notice that $\sqrt{w_1^2 + w_2^2}$ is the Euclidean (L2) norm of the weights. With more than two features, this distance generalizes to $\frac{2}{||w||}$, which is what you learned in class. Therefore, a larger weight vector results in a smaller distance between the two boundaries, aka a smaller margin.