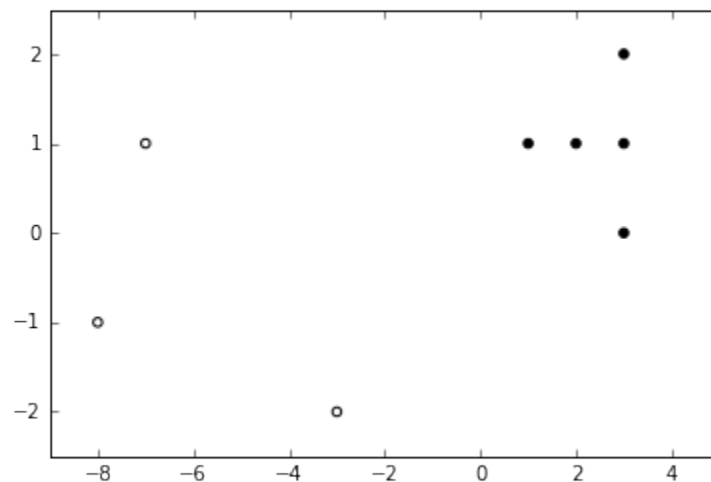


Separating Hyperplanes

We'll soon introduce a new type of classifier, called a *support vector machine*. We'll focus on *linear* support vector machines, in particular a *maximal margin classifier*.

The big idea is to pick a line so that all of the data from one class is on one side of the line, and all of the data from the other class is on the other side of the line. We also want to do this in the way that creates the most separation.

Let's see how this works for some fake data. Draw a line to separate the two groups plotted below.



Compare the line you drew with the lines your neighbors drew. Do you have the same line? How did you decide what line to draw?

Discuss the following questions with your neighbors.

1. If I removed the white point at $(-8, -1)$, would you change your separating line?
2. If I removed the black point at $(3, 2)$, would you change your separating line?
3. If I removed the white point at $(-7, 1)$, would you change your separating line?
4. If I removed the black point at $(3, 0)$, would you change your separating line?
5. Given your answers to the above, which points do you think matter the most?

The points that matter the most are called the *support vectors*.

In order to describe our dividing line, we'll use a linear function, something like $f(x_1, x_2) = 3x_1 + 2x_2 - 1$. The dividing line should be the set of points (x_1, x_2) such that $f(x_1, x_2) = 0$. Then the two sides of the line are where $f(x_1, x_2) > 0$ and where $f(x_1, x_2) < 0$.

Let's see how this works for our fake data.

6. Write down the equation of your line in $y = mx + b$ form.

Now, we'll find our function $f(x_1, x_2)$.

7. Change your x to x_1 , and change your y to x_2 .

8. Subtract everything over the right side of the equation, so the left side is just 0.

9. The function f is defined by the right side of the equation above. Write

$$f(x_1, x_2) = \underline{\hspace{10em}}$$

10. Plug in a few points to your new equation: if you plug in coordinates for white dots, do you get positive or negative numbers? If you plug in coordinates for black dots, do you get positive or negative numbers?

If all the white dots give you outputs with the same sign from $f(x_1, x_2)$, and all the black dots give you outputs with the opposite sign, you have made your first *separating hyperplane*.

We'll use this idea of separating hyperplane to create a classifier called a support vector machine, which probably seems like a strange term to use. Let's break it down.

- The support of a function is the set where the function isn't zero. Here it's really about where our function f is positive and where it is negative, since that determines the classes.
- More specifically, the support vectors are the training data points which are closest to the separating hyperplane. We saw that these points were most important. In this context, a vector is the same as a point: it has coordinates. A two-dimensional vector \mathbf{x} is written (x, y) or (x_1, x_2) , for instance. In other contexts, we might think of a vector as an arrow that points to a point, so it has a direction and a magnitude.
- Machine: it sounds cool! Also, a machine is something designed to accomplish some task. Here, the task is classifying our data points.