## Support Vector Machines

We've seen that we can often draw many different separating hyperplanes to divide our data into classes, and we have some intuitive idea of what creates the largest separation. But how can we do this mathematically?

This is a type of problem called an *optimization* problem. "Optimal" means "best," measured in a specific mathematical way. In this case, we want to find the optimal separating hyperplane.

The people who invented support vector machines (SVM) decided that they wanted the widest possible "street" between the two groups of data, with the separating line running down the middle of the street. This is called the margin. We use the "street" analogy because there will be a lane on each side of the separating line!

Here is a picture from page 342 of the book, "Introduction to Statistical Learning with R" (found at `https://www-bcf.usc.edu/~gareth/ISL/`).[1]
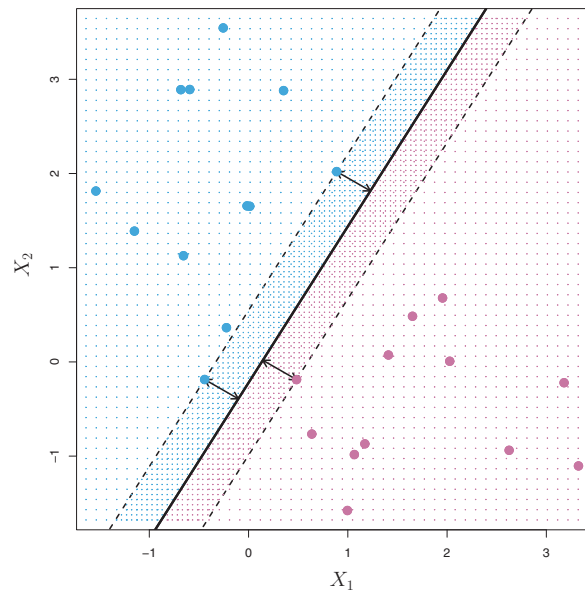
FIGURE 9.3. *There are two classes of observations, shown in blue and in purple. The maximal margin hyperplane is shown as a solid line. The margin is the distance from the solid line to either of the dashed lines. The two blue points and the purple point that lie on the dashed lines are the support vectors, and the distance from those points to the margin is indicated by arrows. The purple and blue grid indicates the decision rule made by a classifier based on this separating hyperplane.*

The goal is to maximize the margin around the separating hyperplane. This optimization turns out to be not that hard if you are in multivariable calculus ☺ but if you are in algebra, trig, or single-variable calculus, it is very hard. Fortunately, python is going to do the work for us!

---

[1]This is a great book!