Read: ISLR 9.1–9.4

Read James et al., Introduction to Statistical Learning (ISL) 9.1–9.4 (pages 337–356) [on the author's website (https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf)](https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf), which you should have downloaded.

Important Points

- **Hyperplanes**: For reasons of easy visualization, many examples we see in lectures and readings concern low-dimensionality data (with many features $f \leq 3$, generally). This means that when we draw separators between data points, we end up with a line (for 2-dimensional data) or a plane (for 3). A hyperplane is just the mathematical generalization of the concept; those more familiar with high-dimensional geometry may be more comfortable with the idea, but for the rest of us, we can think of these things as "line-like" and the linear separator for data of that dimensionality. The complexity of what these things actually "look" like is why, when we expand the dimensionality of 2-dimensional data using a polynomial transform and then plot the regression function on that more complex data back in the 2-D plane, it can look like a curve—it is the projection of a higher-dimensional feature back into lower dimensions. (If this bugs you, think: "it's like a line.")
- **Maximal margin classifiers**: An essential feature of the SVM is that it achieves a maximal margin classification, separating data in a maximally robust way. Other classifiers—like logistic regression—can achieve this, at least under some circumstances, so it is not only the SVM that does this.
- **SVM power**: SVMs have other properties that are of importance, as discussed in the chapter, including the fact that their performance is less sensitive to data outliers than are many other classifiers and because their dual formulation allows us to work with sophisticated kernel functions, producing efficient and robust classifiers.

Things to Keep in Mind

As you read this section, make sure you understand the following:

1. What exactly is meant by a maximal margin classifier?
2. What features of the SVM, in particular, make it more robust in the presence of outliers in data?
3. Why would we use a kernel function like the RBF with a classifier like an SVM?