MA477: Data Science

Lesson 16 Board Sheet — 19 February 2026

United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

K-Nearest Neighbors Lesson Objectives

- Understand the strengths and limitations of k nearest neighbors (p.164 of ISLP).
- Understand how the bias-variance tradeoff applies when selecting k.
- Use modules in sklearn to assess k nearest neighbors classifiers.

Discussion Questions

We will organize this lecture around the following five questions:

1. Define and describe a hyperplane in two dimensional space?
2. How Can a Hyperplane Be Used for Classification?assumptions important?
3. If Many Separating Hyperplanes Exist, Which One Should We Choose?
4. How Do We Express the Margin Mathematically?
5. What Is the Maximal Margin Optimization Problem?

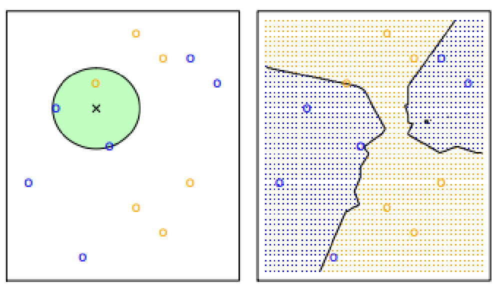

FIGURE 2.14. *The KNN approach, using $K = 3$, is illustrated in a simple situation with six blue observations and six orange observations.* Left: *a test observation at which a predicted class label is desired is shown as a black cross. The three closest points to the test observation are identified, and it is predicted that the test observation belongs to the most commonly-occurring class, in this case blue.* Right: *The KNN decision boundary for this example is shown in black. The blue grid indicates the region in which a test observation will be assigned to the blue class, and the orange grid indicates the region in which it will be assigned to the orange class.*

Figure 1

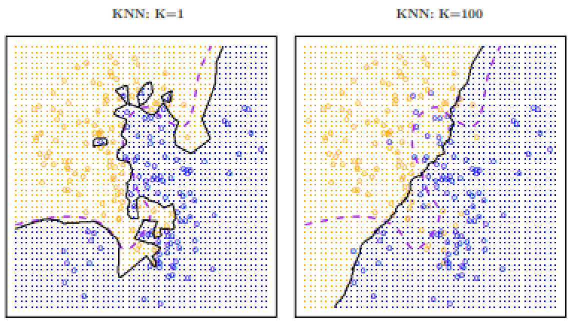

FIGURE 2.16. *A comparison of the KNN decision boundaries (solid black curves) obtained using $K = 1$ and $K = 100$ on the data from Figure 2.13. With $K = 1$, the decision boundary is overly flexible, while with $K = 100$ it is not sufficiently flexible. The Bayes decision boundary is shown as a purple dashed line.*

Figure 2

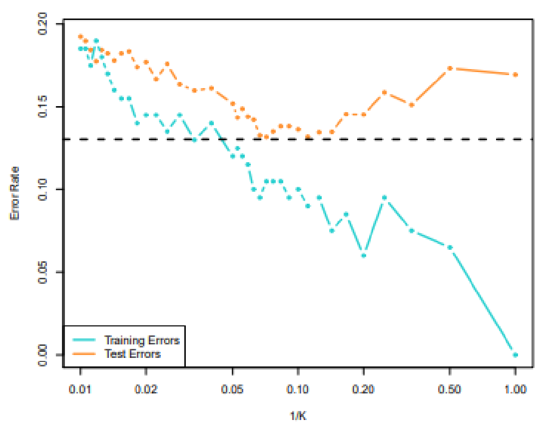

FIGURE 2.17. *The KNN training error rate (blue, 200 observations) and test error rate (orange, 5,000 observations) on the data from Figure 2.13, as the level of flexibility (assessed using $1/K$ on the log scale) increases, or equivalently as the number of neighbors K decreases. The black dashed line indicates the Bayes error rate. The jumpiness of the curves is due to the small size of the training data set.*

Figure 3