MA477: Data Science

Lesson 17 Board Sheet — 24 February 2026

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

1 Classification Comparison Lesson Objectives

- Review all previous classification lessons since lesson 9

Discussion Questions

1. In a soft-margin SVM, how does the parameter C control the tradeoff between a wide margin and allowing margin violations?
2. What does it mean to *maximize the margin* in an SVM? Why can a larger margin improve generalization to new data?
3. What is the main purpose of a kernel function in SVM? Explain what it means to *implicitly* work in a higher-dimensional feature space.
4. In an RBF-kernel SVM, what does γ control? Describe how very large versus very small γ affects the decision boundary and the bias–variance tradeoff.
5. Why is feature standardization especially important for distance-based or similarity-based methods? Name two model families where scaling is typically critical and explain why.
6. How does the choice of k change KNN behavior? Describe what happens at $k = 1$ and what happens as k becomes very large, in terms of bias and variance.
7. Why can KNN perform poorly in high-dimensional feature spaces? Explain why distances become less informative as dimension increases.
8. Conceptually, how does logistic regression differ from an SVM classifier? Discuss probability modeling versus margin maximization, and how this relates to interpretability.

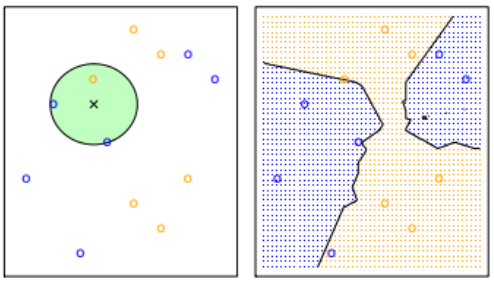

FIGURE 2.14. *The KNN approach, using $K = 3$, is illustrated in a simple situation with six blue observations and six orange observations.* Left: *a test observation at which a predicted class label is desired is shown as a black cross. The three closest points to the test observation are identified, 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

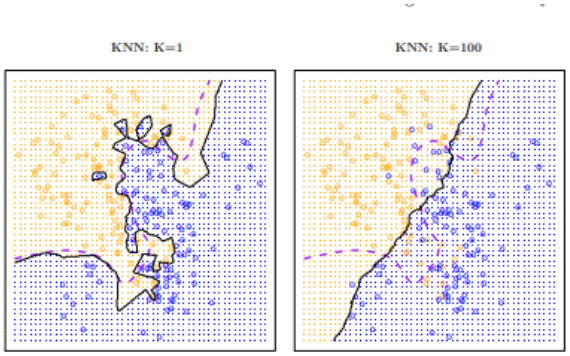

FIGURE 2.16. *A comparison of the KNN decision boundaries (solid black curves) obtained using $K = 1$ and $K = 100$ on the data from Figure 2.13. With $K = 1$, the decision boundary is overly flexible, while with $K = 100$ it is not sufficiently flexible. The Bayes decision boundary is shown as a purple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

Figure 2

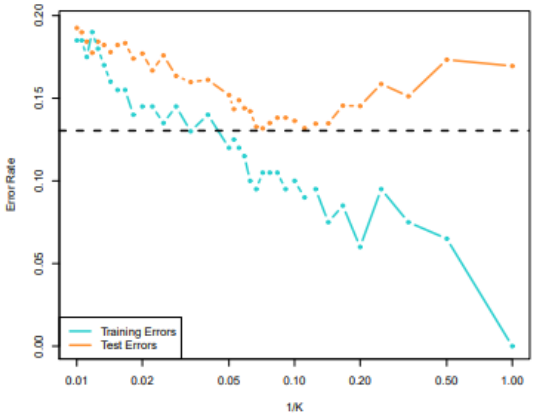

FIGURE 2.17. *The KNN training error rate (blue, 200 observations) and test error rate (orange, 5,000 observations) on the data from Figure 2.13, as the level of flexibility (assessed using $1/K$ on the log scale) increases, or equivalently as the number of neighbors K decreases. The black dashed line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

Figure 3