K-Nearest Neighbors

Forrest Davis

September 2, 2025

Fall 2025 COSC410A: Applied Machine Learning
Colgate University

Example Data

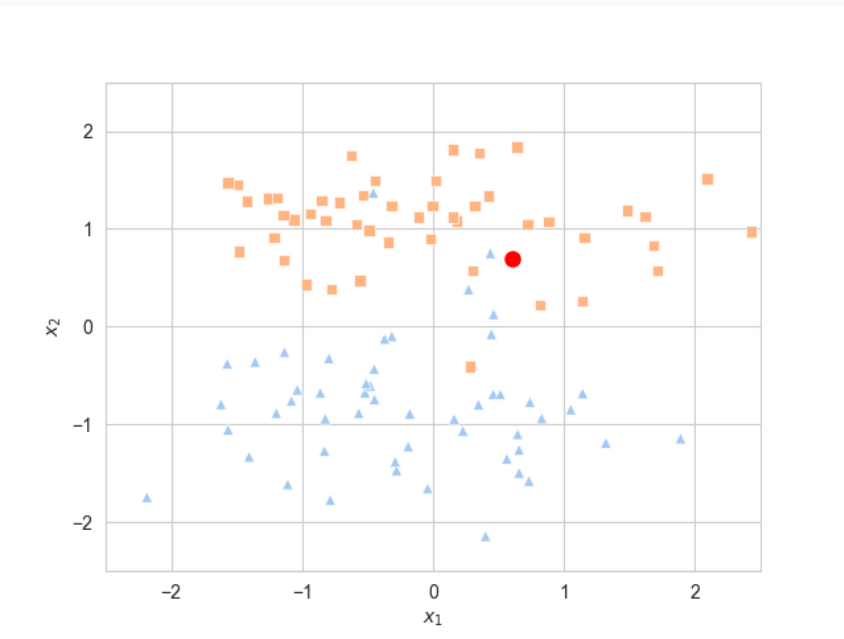

Figure 1: Data with two labels, triangles and squares. One additional data point has been added as a red circle.

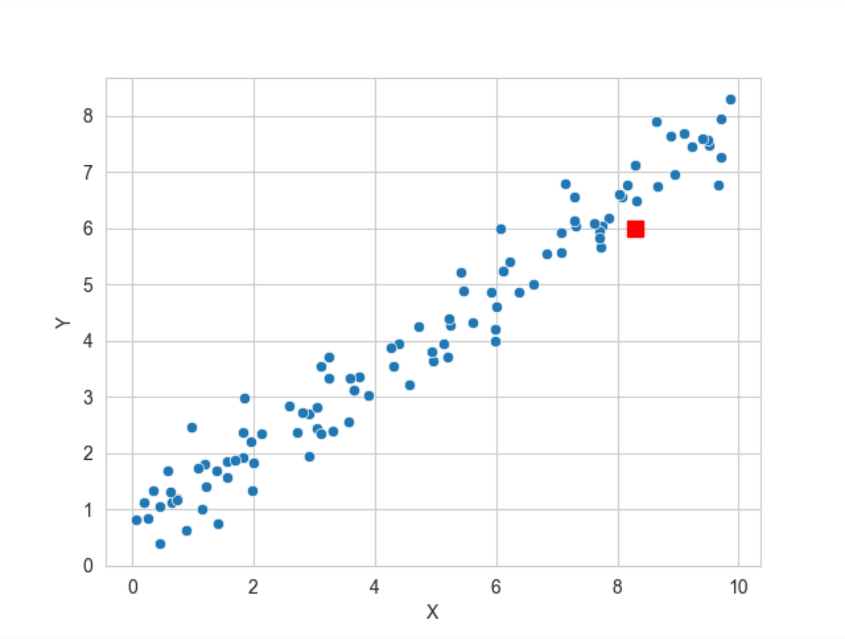

Figure 2: Data with one feature (X) and one label (continuous; Y). One additional data point has been added as a red square.

Distance Metrics

$$\text{euclidean} = \sqrt{\sum_{i=0}^{n-1}(p_i - q_i)^2} \quad : \quad p, q \in \mathbb{R}^n$$

$$\text{manhattan} = \sum_{i=0}^{n-1}|p_i - q_i| \quad : \quad p, q \in \mathbb{R}^n$$

Example Data

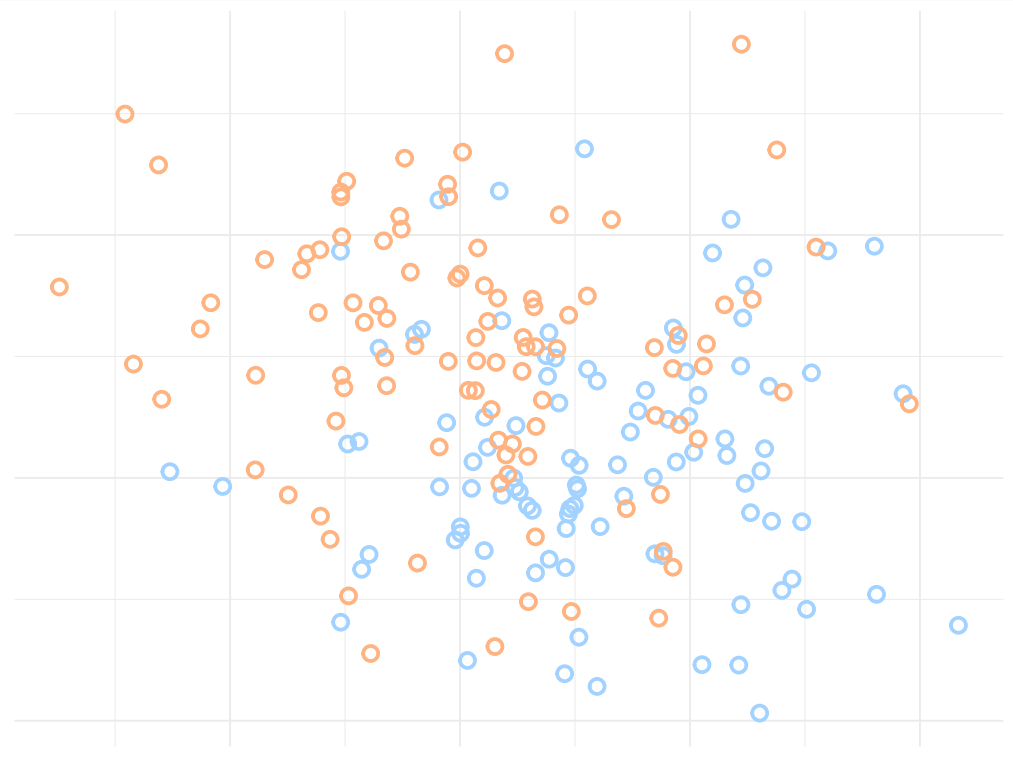

Figure 3: Data with two features and two possible labels (orange and blue)

Fitting Data with varying k

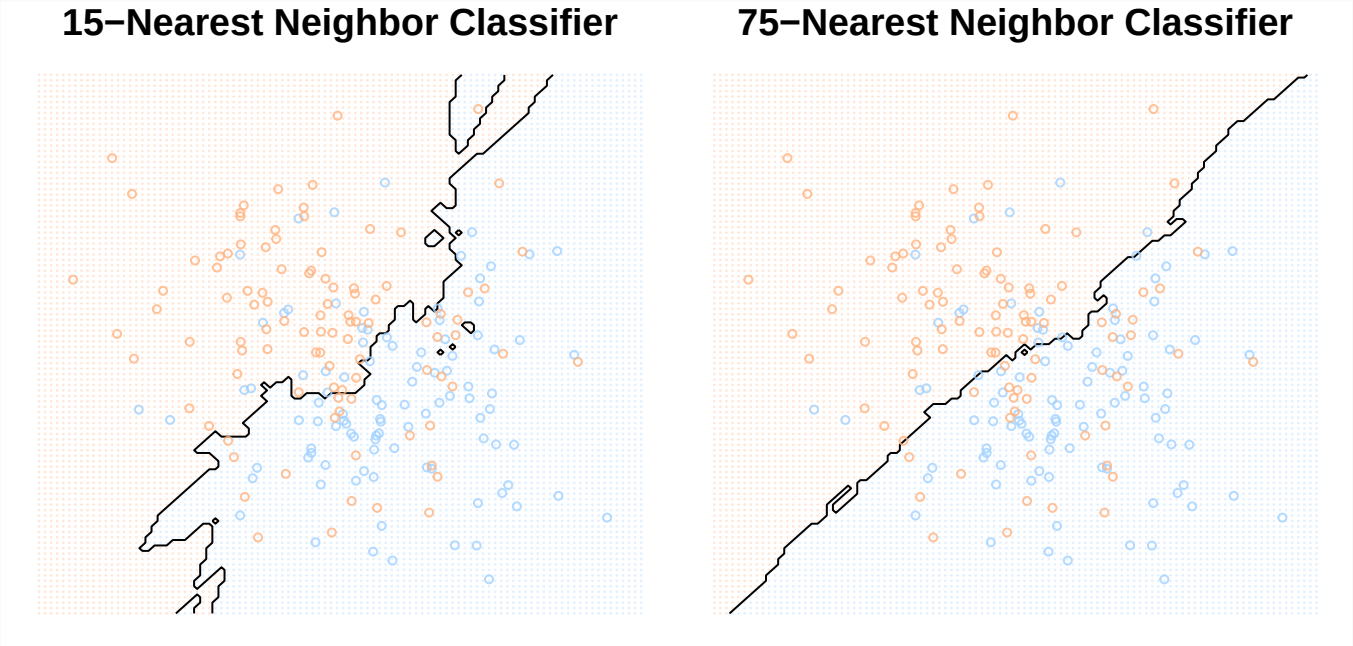

Figure 4: Decision boundary (in black) when fitting two nearest neighbor models to the data in Figure 3.

Fitting Data with varying k

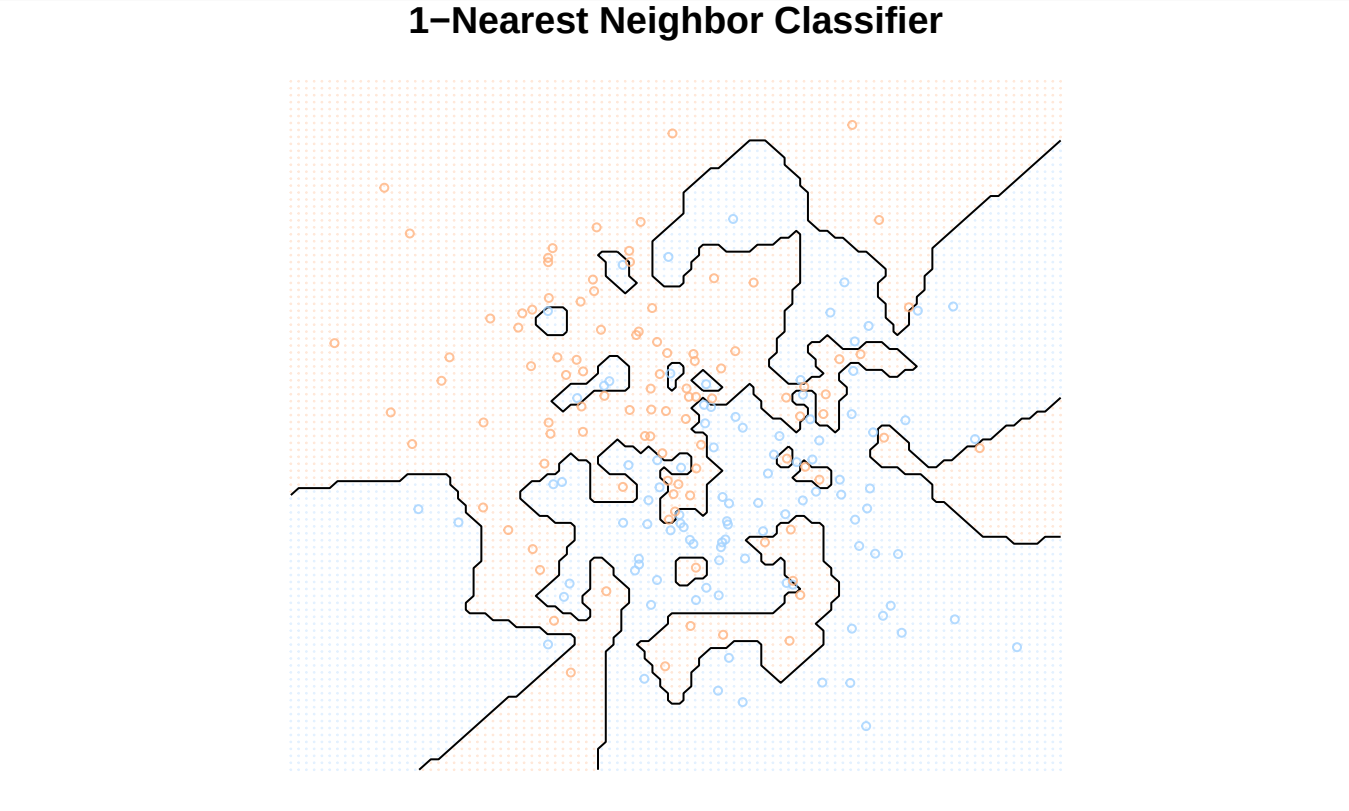

Figure 5: Decision boundary (in black) when fitting the data in Figure 3 with a k of 1.

Varying k and Error Rates

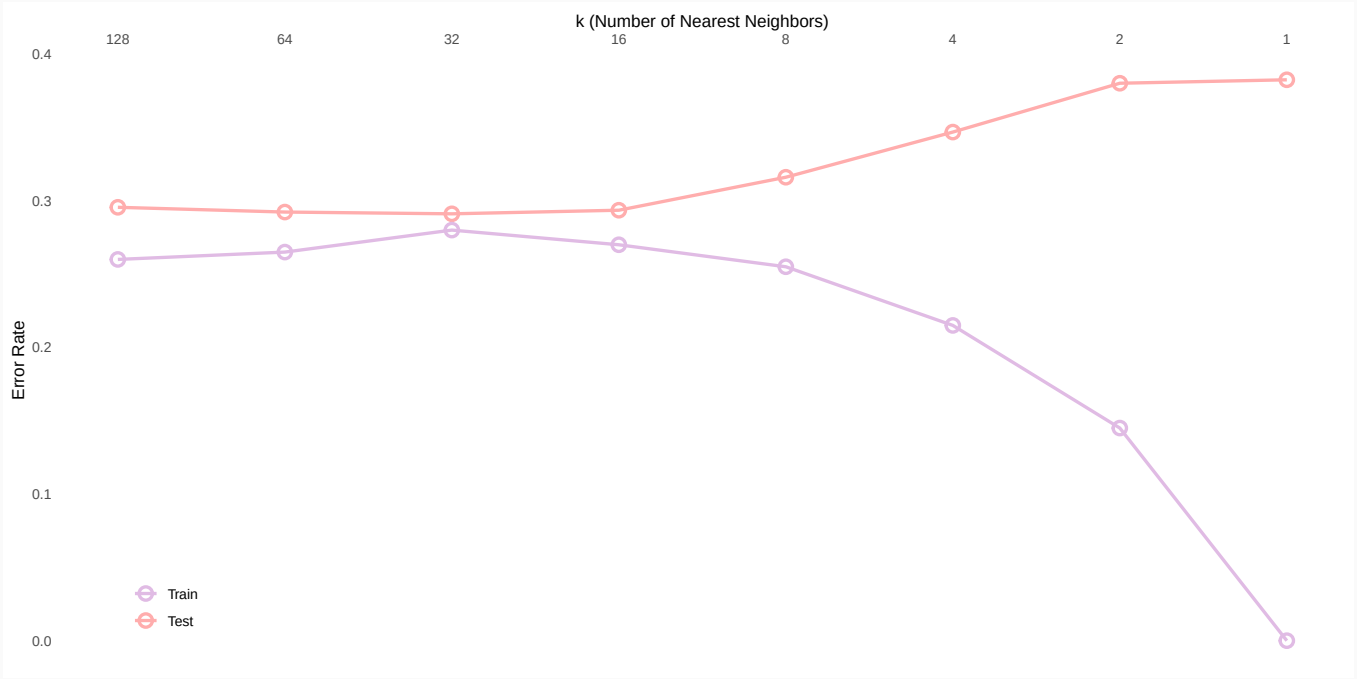

Figure 6: Error rate (y-axis) as a function of k (x-axis). A model trained on 200 samples of data like that in Figure 3 and tested on 10,000 further samples. The line on top (in red) is the error rate on held out data (the 10,000 further samples). The line on the bottom (in purple) is the error rate on the training data.

Effect of Scaling Data

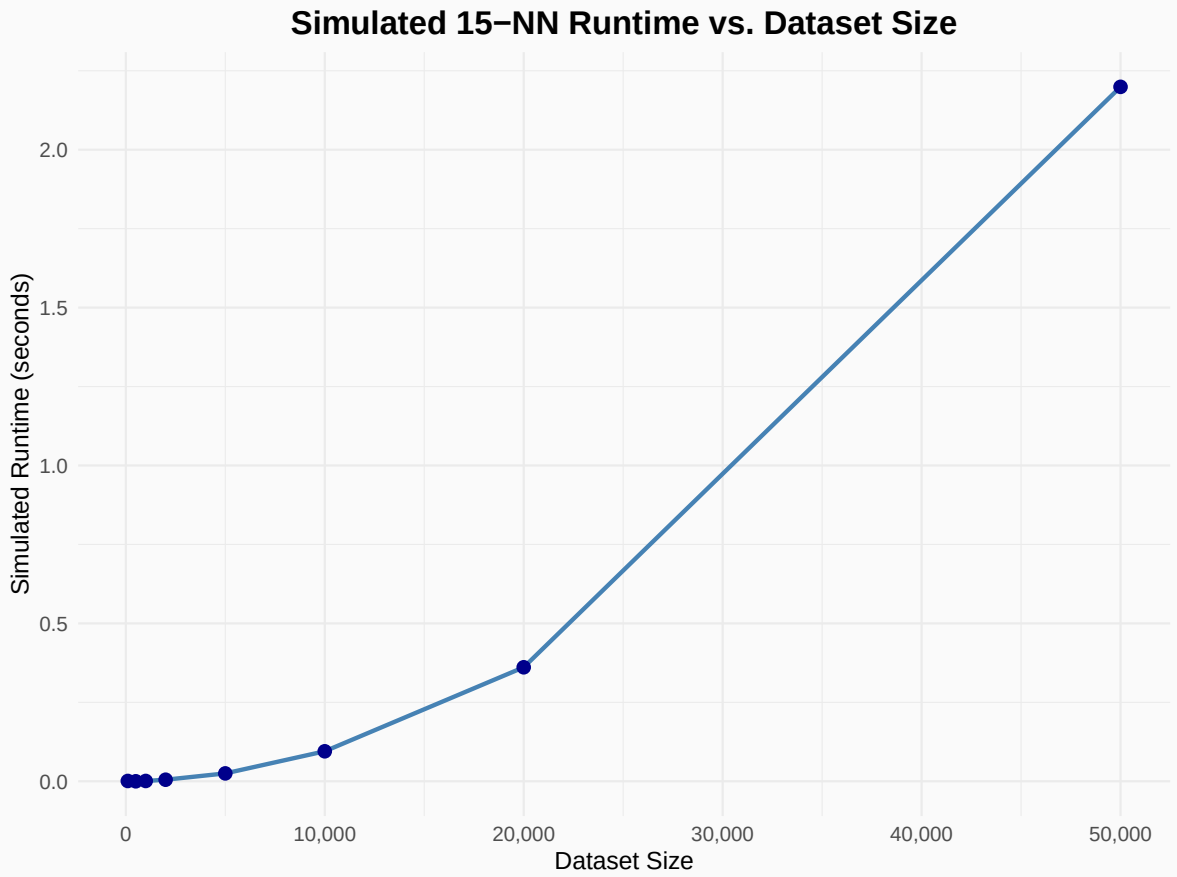

Figure 7: Runtime (y-axis) as a function of dataset size (x-axis) for 15-nearest neighbors.

Influence of Distant Neighbors

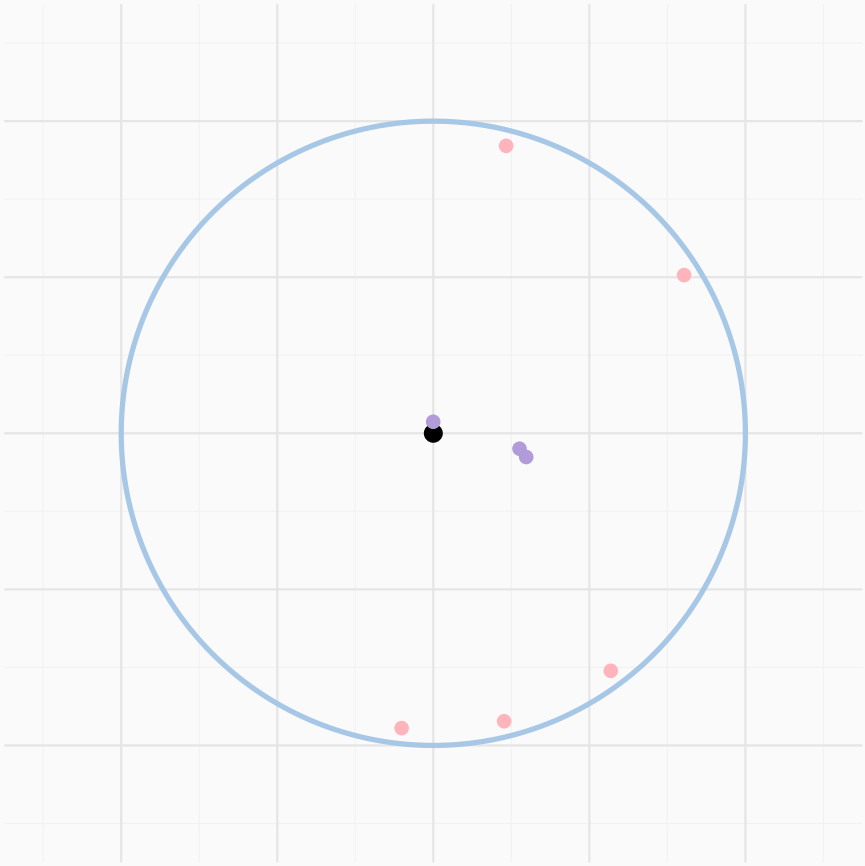

Figure 8: Sample data point (in black) and the 8 nearest neighbors, 3 are close (in purple) and 5 are far (in red). The circle denotes the maximum distance for this neighborhood.

Features of Different Scales

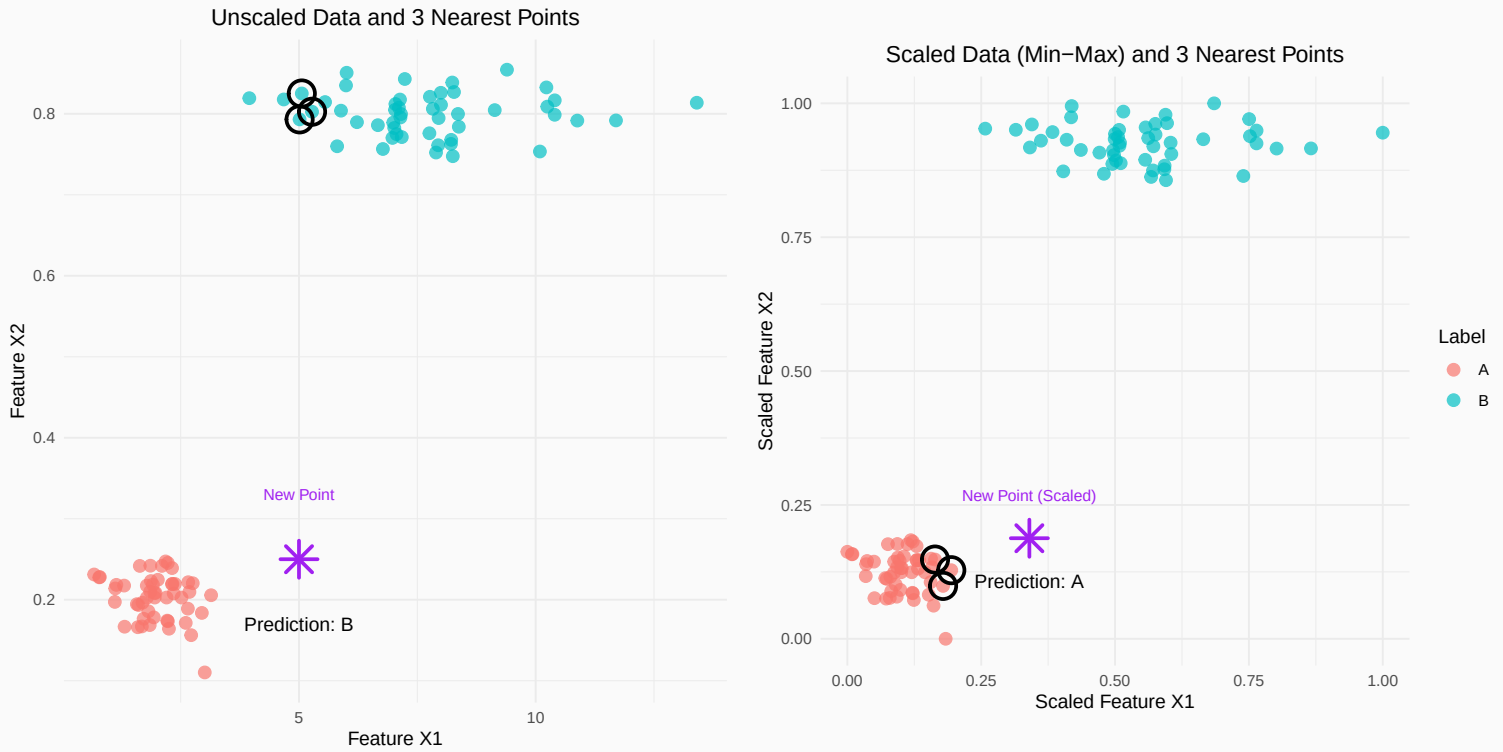

Figure 9: Effect on feature scale on classification of a new point (in purple) in one of two categories (red and blue). On the left is the raw data. On the right is the same data but with normalization so they have the same range.

Curse of Dimensionality

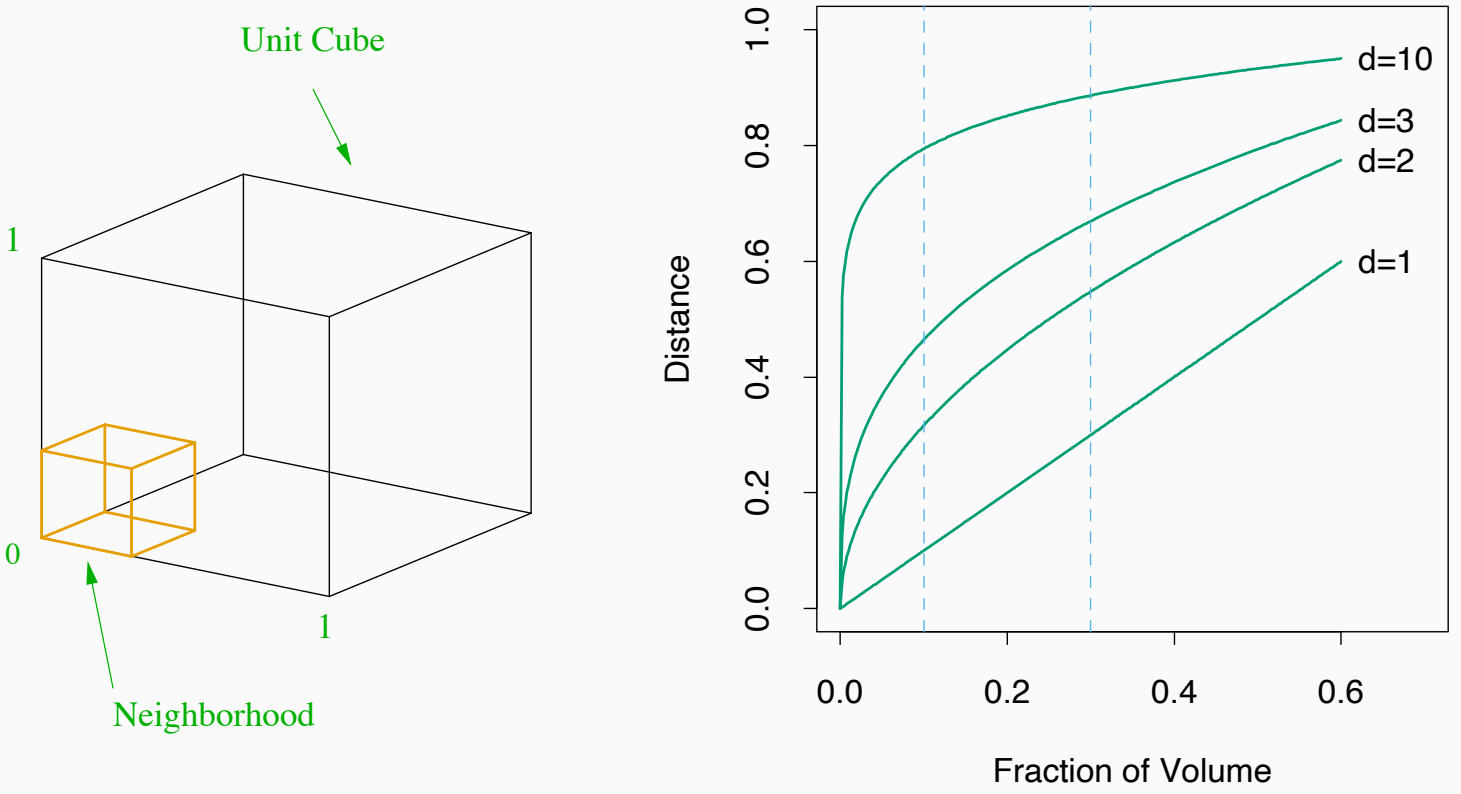

Figure 10: Illustration (corresponding to Figure 2.6 in Hastie, 2001) of the curse of dimensionality. On the left, we have a subcubical neighborhood for uniform data in a unit cube. The figure on the right shows the side-length of the subcube needed to capture a fraction of the volume of the data, for different dimensions.

References

Hastie, T. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York.