Nearest Neighbor Predictors
Outline

1. Nearest Neighbor Prediction
2. Complexity Considerations
3. The Voronoi Diagram
4. Overfitting and $k$ Nearest Neighbors
Nearest Neighbor Prediction

- NN is very simple: This is why we start here
- NN is very unusual:
  - No training!
  - Slow inference (using the predictor)
  - $Y$ can be *anything*
  - Almost no difference between regression and classification
  - Hypothesis space hard to define
How it Works

• Given $T = \{(x_1, y_1), \ldots, (x_N, y_N)\}$

• Just store $T$ (memorization)

• Need a distance in the data space $X$

• Perhaps $\Delta(x, x') = \|x - x'\|^2$

• Then, $h(x) = y_{\nu(x)}$

  where $\nu(x) \in \arg \min_{n=1,\ldots,N} \Delta(x, x_n)$

• Return the value $y_n$ for the training point $x_n$ that is nearest to $x$
How to find $\nu(x)$?

$\nu(x) = \arg \min_{n=1,\ldots,N} \Delta(x, x_n)$

- Compute all $\Delta(x, x_n)$ and find the smallest
- $O(Nd)$ (where $x \in \mathbb{R}^d$)
- Cannot do better exactly
- Can do better if we accept $\Delta(x, x_{\nu(x)}) < (1 + \epsilon)\Delta(x, x_{\nu^*(x)})$ for some $\epsilon > 0$
- “Approximate NN” uses $k$-$d$ trees, R-trees, locality sensitive hashing
The Voronoi Diagram

- Only conceptual, or for $d = 2, 3, \text{maybe 4}$
- $\Theta(N \log N + N^{\lceil d/2 \rceil})$
Decision Boundary
Overfitting
$k$ Nearest Neighbors

- Retrieve the $k$ nearest neighbors $x_1, \ldots, x_k$ of $x$
- Return a *summary* of the corresponding $y_1, \ldots, y_k$
- Classification summary: majority
- Regression summary: Mean, median
Less Overfitting ($k = 9$)
A Simple Regression Example, $\mathbb{R} \rightarrow \mathbb{R}$