CSCI 6521

Advance Machine Learning I

Supp. Material:

kNN & Curse of Dim.

Md Tamjidul Hoque

Simple Approach 2: Nearest Neighbors or kNN:

kNN Applied to Housing dataset

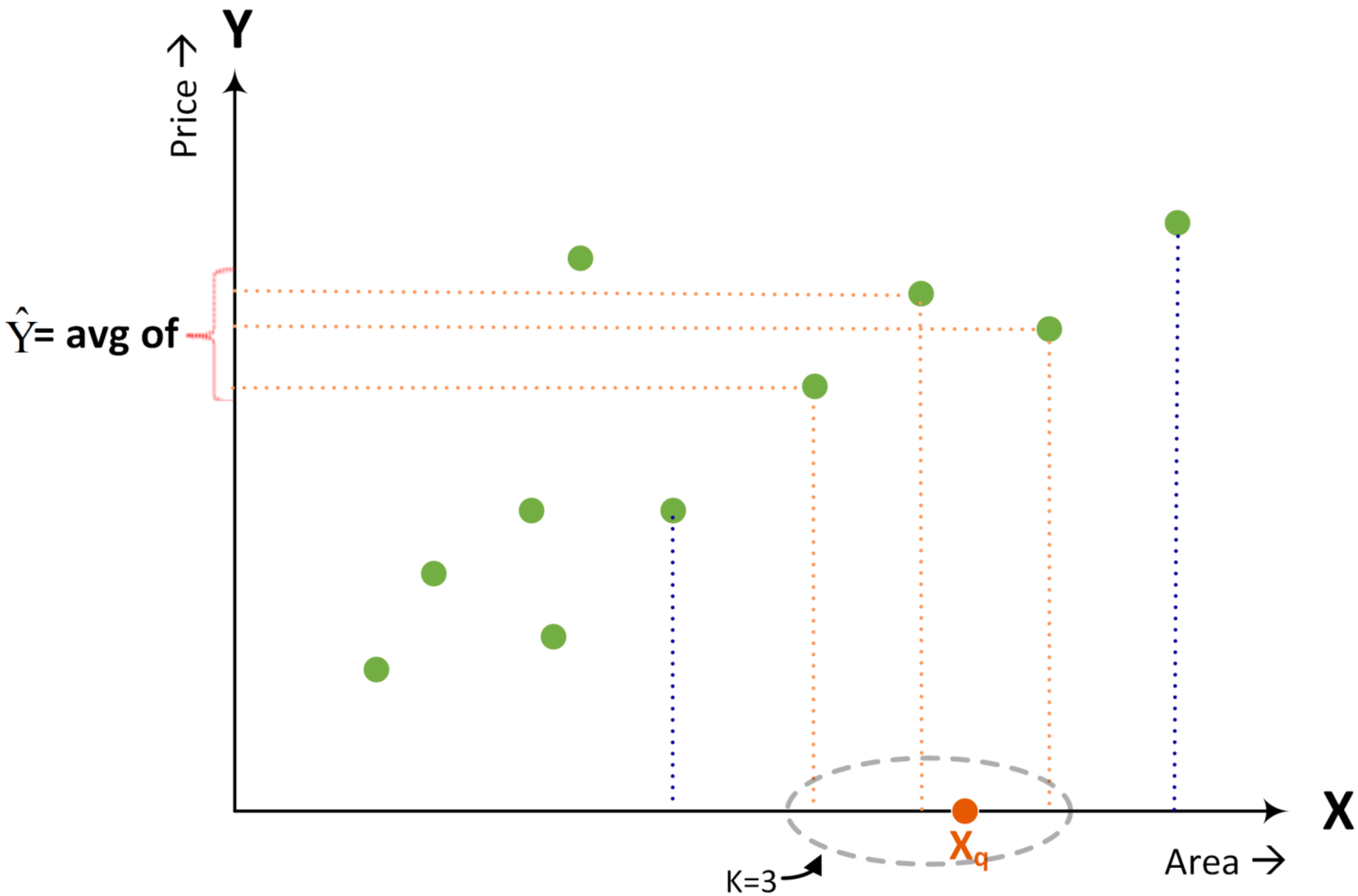


Figure: Area versus house price. Schematic outline of predicting the price for area x_q using the k-nearest neighbor method when k=3

…Simple Approach 2: Nearest Neighbors

- Idea: Use those observations in the training set that are closest to the given input

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \qquad (2.8)$$

  - $N_k(x)$ is the set of the k closest points to x is the training sample.
  - Average the outcome of the k closest training sample points
  - Fewer training points are misclassified.

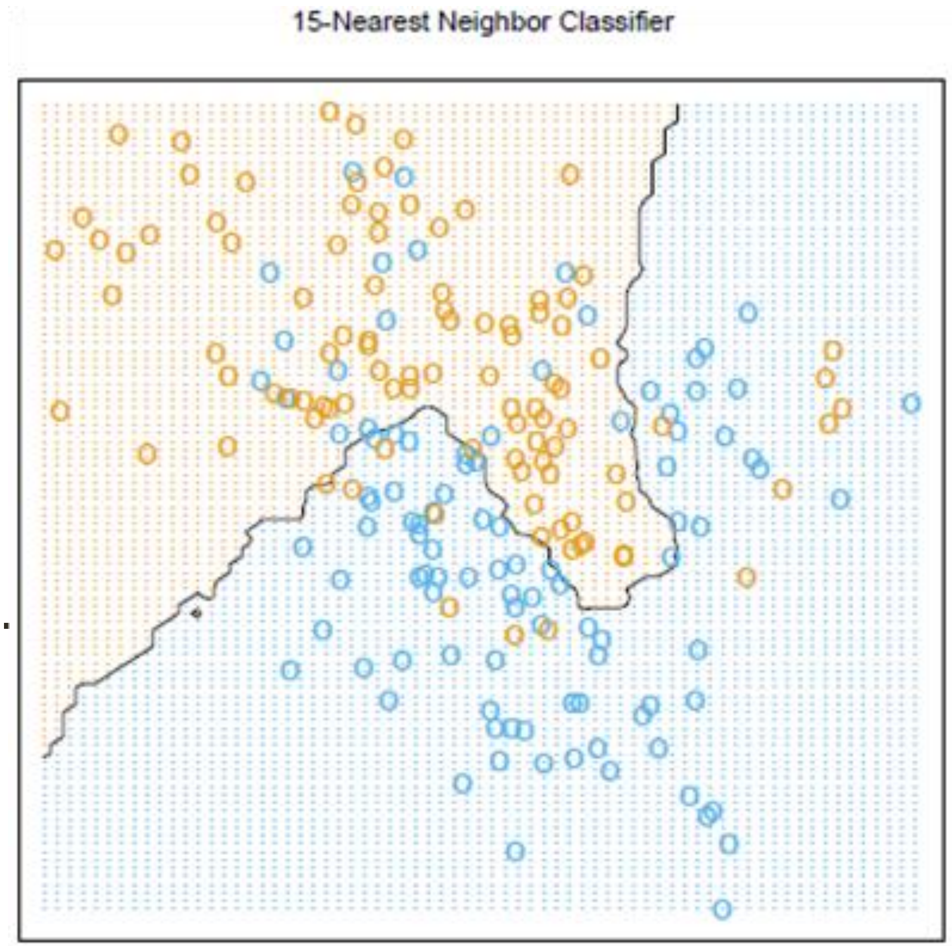
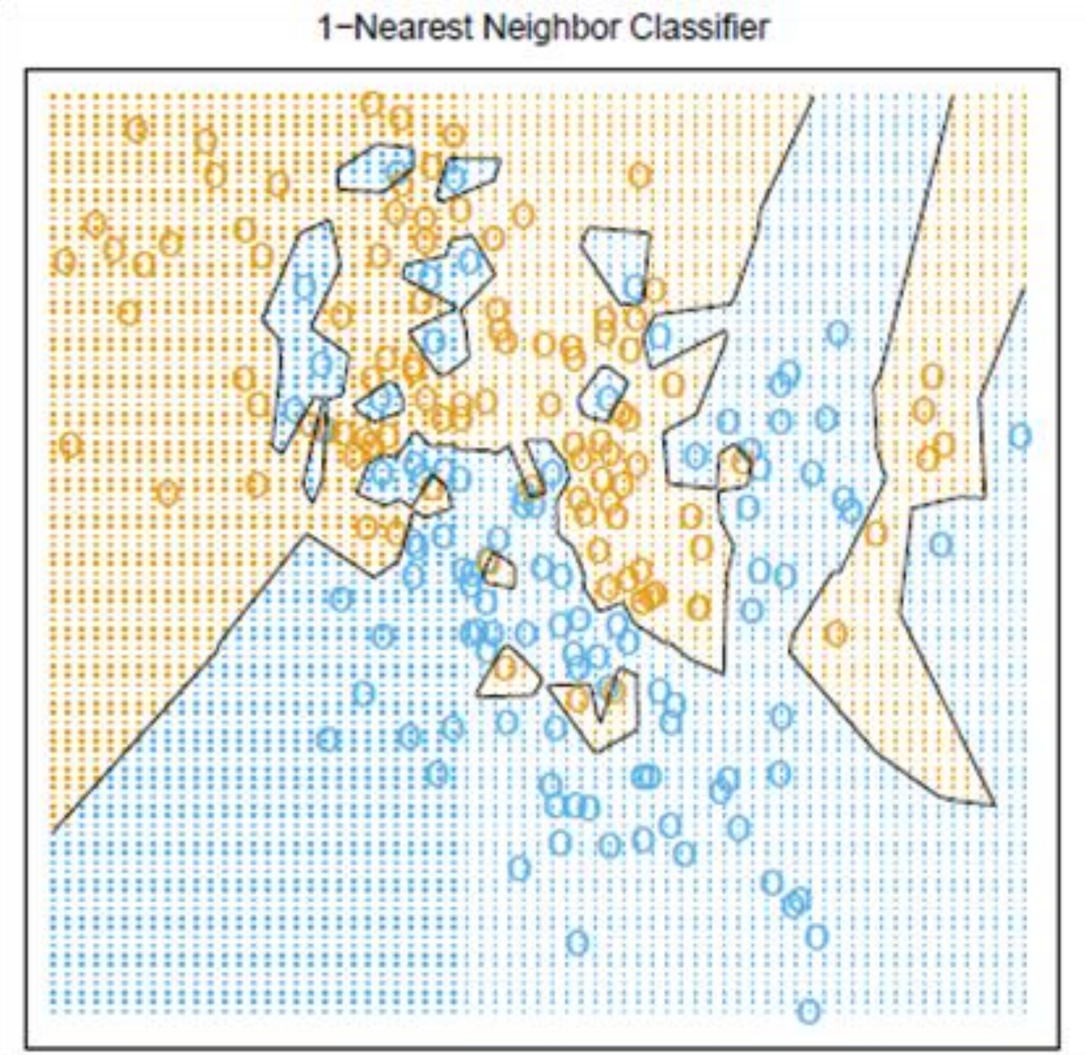


Figure 2.2. *The classes are coded as a binary variable* (BLUE = 0, ORANGE = 1) *and then fit by* 15*-nearest-neighbor averaging as in equation* (2.8)*. The predicted class is hence chosen by majority vote amongst the* 15*-nearest neighbors*

Simple Approach 2: Nearest Neighbors



- For 1-NN, there is zero misclassification.

Figure 2.3. *The classes are coded as a binary variable* (BLUE = 0, ORANGE = 1), *and then predicted by* 1-*nearest-neighbor classification*

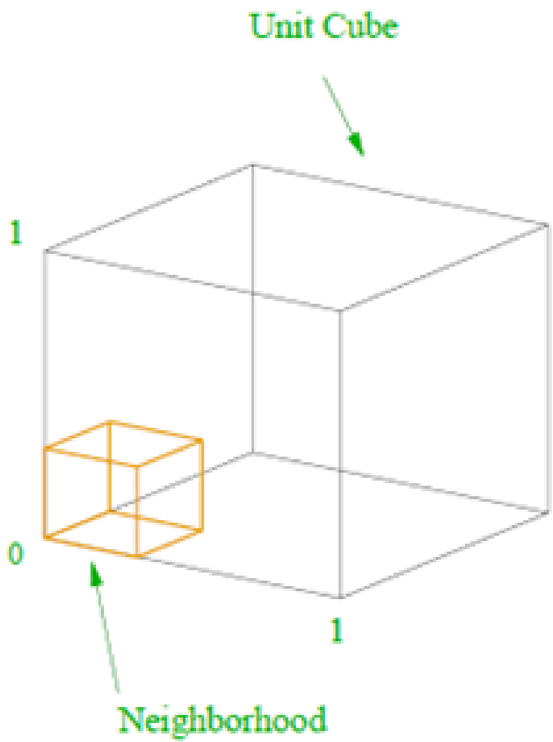
Curse of Dimensionality / Local Methods in High Dimensions

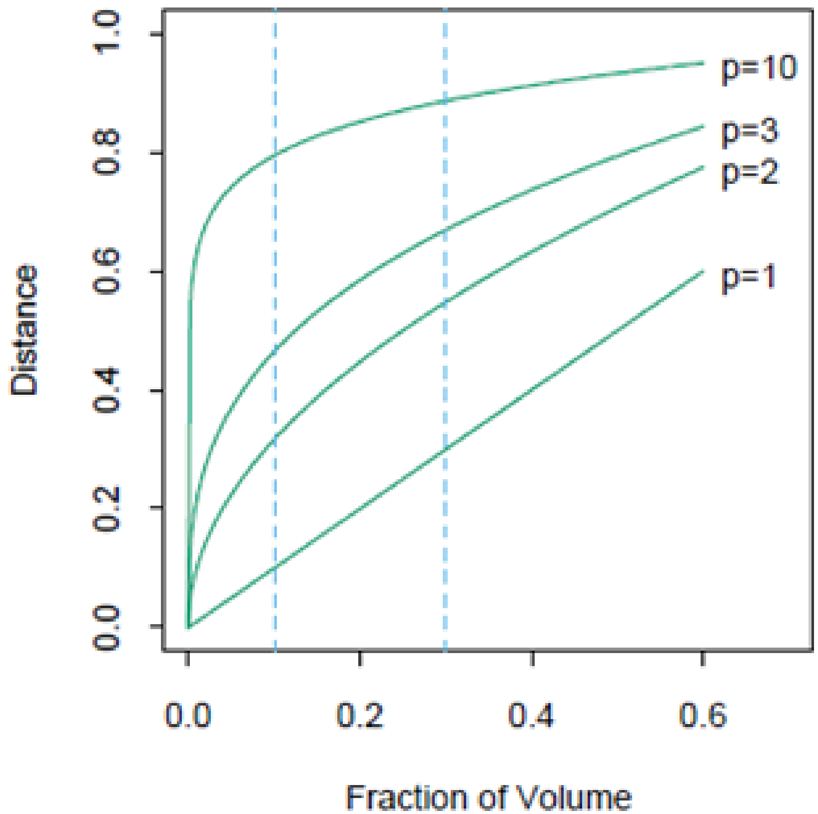
- So far we have seen 2 extremes:
  - the stable but biased **linear model** and
  - the less stable but apparently less biased class of ***k*-NN** estimates.

- It would seem that with a reasonably large set of training data, we could always approximate the theoretically optimal conditional expectation by k-NN averaging, since we should be able to find a fairly large neighborhood of observations close to any *x* and average them.

- This approach and our intuition breaks down in high dimensions, and the phenomenon is commonly referred to as the *curse of dimensionality.* Let us see, how!?

Curse of Dimensionality / Local Methods in High Dimensions





- Consider the nearest-neighbor procedure for inputs uniformly distributed in a p-dimensional unit hypercube.

- Suppose we send out a hypercubical neighborhood about a target point to capture a fraction r of the observations.

Since this corresponds to a fraction r of the unit volume, the expected edge length will be $e_p(r) = r^{1/p}$.

Curse of Dimensionality / Local Methods in High Dimensions

- In ten dimensions $e_{10}(0.01) = 0.63$ and $e_{10}(0.1) = 0.80$, while the entire range for each input is only 1.0.
- So, to capture 1% or 10% of the data to form a local average, we must cover 63% or 80% of the range of each input variable, respectively.
- Such neighborhoods are no longer “local.”
- Reducing *r* dramatically does not help much either, since the fewer observations we average, the higher is the variance of our fit.

Curse of Dimensionality and Statistical Model

- If the dimension of the input space is high, the nearest neighbors need not be close to the target point, and can result in large errors;

- If special structure is known to exist, this can be used to reduce both the bias and the variance of the estimates.