STAT 745 – Fall 2014
Assignment 2

Group 1: Francene Cicia, Doug Raffle, Melissa Smith

September 3, 2014

1 Least Squares vs. k-NN Classifiers

Let $\mathbf{y}$ be the response and $\mathbf{X}$ the data such that:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1-\tau & \tau \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix}, \quad \tau \in \{0, 1\}$$

1.a Determine $\hat{\beta}^{(ls)}$

$$\begin{aligned} \hat{\beta}^{(ls)} &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \\ &= \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1-\tau \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & \tau \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1-\tau & \tau \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1-\tau \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & \tau \end{bmatrix} \mathbf{y} \end{aligned}$$

In the case $\tau = 0$:

$$\begin{aligned} \hat{\beta}^{(ls)} &= \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \mathbf{y} \\ &= \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \mathbf{y} \\ &= \begin{bmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} \mathbf{y} \end{aligned}$$

$$\hat{\beta}^{(ls)} = \begin{bmatrix} \frac{1}{5}\left(y_8 + \sum_{i=1}^{4} y_i\right) \\ \\ \frac{1}{3}\sum_{i=5}^{7} y_i \end{bmatrix}$$

In the case $\tau = 1$:

$$\begin{aligned} \hat{\beta}^{(ls)} &= \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \mathbf{y} \\ &= \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \mathbf{y} \\ &= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \mathbf{y} \end{aligned}$$

$$\hat{\beta}^{(ls)} = \begin{bmatrix} \frac{1}{4}\sum_{i=1}^{4} y_i \\ \\ \frac{1}{4}\sum_{i=5}^{8} y_i \end{bmatrix}$$

Leaving us with the final expression:

$$\hat{\beta}^{(ls)} = \begin{cases} \begin{bmatrix} \frac{1}{5}\left(y_8 + \sum_{i=1}^{4} y_i\right) \\ \\ \frac{1}{3}\sum_{i=5}^{7} y_i \end{bmatrix} & \text{if } \tau = 0 \\ \\ \begin{bmatrix} \frac{1}{4}\sum_{i=1}^{4} y_i \\ \\ \frac{1}{4}\sum_{i=5}^{8} y_i \end{bmatrix} & \text{if } \tau = 1 \end{cases}$$

1.b k-NN Classifier

$$\hat{y} = \frac{1}{|N_k(x)|} \sum_{x \in N_k(x)} y_i$$

In this orthogonal data, the $k = 2$ closest points will be within one group or the other, as $x_i^T \in \{[1 \quad 0], [0 \quad 1]\} \forall x_i^T \in \mathbf{X}$. The data naturally forms two groups: where $x_i^T = [1 \quad 0]$ and where $x_i^T = [0 \quad 1]$, and distances from any given x to the x_i^T's within these groups are identical (tied). Since the group sizes are all greater than k, there is no fair way to determine only two y_i's to average for the classifier without losing information. A reasonable approach would be to expand the neighborhood to include the entire group.

For $\tau = 0$:

$$N_k(x) = \begin{cases} \{x_1^T, x_2^T, x_3^T, x_4^T, x_8^T\} & \text{if } x = [1 \quad 0] \\ \{x_5^T, x_6^T, x_7^T\} & \text{if } x = [0 \quad 1] \end{cases}$$

$$\hat{y} = \begin{cases} \frac{1}{5}\left(y_8 + \sum_{i=1}^{4} y_i\right) & \text{if } x = [1 \quad 0] \\ \frac{1}{3}\sum_{i=5}^{7} y_i & \text{if } x = [0 \quad 1] \end{cases}$$

For $\tau = 1$:

$$N_k(x) = \begin{cases} \{x_1^T, x_2^T, x_3^T, x_4^T\} & \text{if } x = [1 \quad 0] \\ \{x_5^T, x_6^T, x_7^T, x_8^T\} & \text{if } x = [0 \quad 1] \end{cases}$$

$$\hat{y} = \begin{cases} \frac{1}{4}\sum_{i=1}^{4} y_i & \text{if } x = [1 \quad 0] \\ \frac{1}{4}\sum_{i=5}^{8} y_i & \text{if } x = [0 \quad 1] \end{cases}$$

1.c Compare

Both methods would produce the same predictions. For orthogonal data in Least Square Regression without an intercept, the predictions will always be the group means. As discussed above, we are dealing with tied distances by increasing the size of the neighborhoods to include all of the tied x_i^T's, which means our k-NN classifier would again be the mean of the y_i's in the group x belongs to.

For example, consider an observation $x = [1 \quad 0]$ and $\tau = 0$:

Least Squares:

$$\hat{y} = [1 \quad 0] \begin{bmatrix} \frac{1}{5}\left(y_8 + \sum_{i=1}^{4} y_i\right) \\ \\ \frac{1}{3}\sum_{i=5}^{7} y_i \end{bmatrix} = \frac{1}{5}\left(y_8 + \sum_{i=1}^{4} y_i\right)$$

k-NN:

$$N_k(x) = \{x_1^T, x_2^T, x_3^T, x_4^T, x_8^T\}$$

$$\hat{y} = \frac{1}{5}\left(y_8 + \sum_{i=1}^{4} y_i\right)$$

This relationship would hold true for $k = 2$ and $k = 3$ (and even $k = 4$ if $\tau = 1$). As long as k doesn't exceed the size of the smallest group, the two methods will give the same predictions for orthogonal data (assuming we use the same method for handling tied distances).