Math 5364 Homework 7

1. Use R to plot the triweight and cosine kernel functions from Hechenbichler and Schliep (2004). Use a plot window of xlim=c(-1.2,1.2) and ylim=c(0,1.2). (Hint: Boolean commands like (x <= 20) return TRUE/FALSE values, but they are treated as 1’s and 0’s when multiplied by numbers.)

2. Returning to the wdbc.data data set, use train.kknn to find the optimal kernel function and value of $k$ for predicting breast cancer diagnosis using weighted $k$-nearest neighbors.

3. Divide the data into 70% training and 30% testing data, and calculate the test accuracy using the optimal kernel function and value of $k$. Find a 95% confidence interval for the accuracy.

4. For the same training and test data, find the test accuracy using a rectangular kernel and the optimal value of $k$ obtained in homework 6.

5. Test whether the difference in test accuracies is statistically significant.