**Real-Time Location System Case Study**

The assignment for the Real Time Location System (RTLS) unit is a case study using the data that is available for the Nolan and Lang textbook website:

<http://rdatasciencecases.org/Data/offline.final.trace.txt>

An example of one (the first) line of the data within the file is shown below:

t=1139643118358;id=00:02:2D:21:0F:33;pos=0.0,0.0,0.0;degree=0.0;

00:14:bf:b1:97:8a=-38,2437000000,3;

00:14:bf:b1:97:90=-56,2427000000,3;

00:0f:a3:39:e1:c0=-53,2462000000,3;

00:14:bf:b1:97:8d=-65,2442000000,3;

00:14:bf:b1:97:81=-65,2422000000,3;

00:14:bf:3b:c7:c6=-66,2432000000,3;

00:0f:a3:39:dd:cd=-75,2412000000,3;

00:0f:a3:39:e0:4b=-78,2462000000,3;

00:0f:a3:39:e2:10=-87,2437000000,3;

02:64:fb:68:52:e6=-88,2447000000,1;

02:00:42:55:31:00=-84,2457000000,1

This is the first line of the data file, and the components of the line are organized as shown in Table 1, page 7, of the Nolan and Lang book and described below. The variable t indicates the timestamp of the data being gathered. The timestamp is in units of milliseconds and represents the number of milliseconds since midnight, January 1, 1970 UTC. The variable id indicates the MAC address of the scanning device. The variable pos indicates the physical coordinates of the scanning device. The variable degree indicates the orientation of the user carrying the scanning device in degrees. The remainder of the data contains a quadruplet consisting of the MAC address of a responding peer with its corresponding values for received signal strength in dBm, the channel frequency and the devices mode of operation (either 3 for access point or 1 for device in adhoc mode).

For the case study for this unit, we will be analyzing this data using the k-nearest neighbors to determine locations and to determine potential issues with decisions made regarding the use, and non-use, of the data. Section 1.5 of Nolan and Lang provides a basic k-nearest neighbors approach to determining location assuming the floor plan for the building (see Figure 1.1) is accurate. The floor plan shows six access points; however, the data contains seven access points with roughly the expected number of signals. In the analysis presented in Nolan and Lang, the access points were matched to their locations, and the decision was made to keep the access point with MAC address 00:0f:a3:39:e1:c0 and to eliminate the data corresponding to MAC address 00:0f:a3:39:dd:cd.

Conduct a more thorough data analysis into these two MAC addresses including determining locations by using data corresponding to both MAC addresses. Which of these two MAC addresses should be used and which should not be used for RTLS? Which MAC address yields the best prediction of location? Does using data for both MAC addresses simultaneously yield more, or less, accurate prediction of location? (Note: this portion is derived from Exercise Q.9 in Nolan and Lang.)

While k-nearest neighbors has proven to be a good approach to determining location, alternate approaches have been proposed. One simple alternative approach is to use weights on the received signal strength, where the weight is inversely proportional to the “distance” from the test observation. This allows for the “nearest” points to have a greater contribution to the k-nearest neighbor location calculation than the points that are “further” away.

Implement this alternative prediction method. For what range of values of weights are you able to obtain better prediction values than for the unweighted k-nearest neighbor approach? Use calcError() to compare this approach to the simple average.

Create an iPython Notebook including code output and graphics for all of your work.

Include an introduction to explain the case study, explain the approach used to complete the case study and explain the output achieved. Explanations of output should be included as close to the output or figures as possible.

List all references used, including the book by Nolan and Lang.