

1 Background

At home, your computer connects to the internet by first going through a wireless link to your WiFi router, and then through a series of wired links via your service provider to the internet. In large WiFi networks, the task of connecting the wireless links with the wired links is carried out by many wireless access points (APs) that coordinate with each other, thus covering a larger area. Every client device periodically discovers the list of nearby APs by exchanging probing packets. This allows the client device to select the best AP (with the strongest probe response signal) with which to establish the wireless link. The signal strength (in decibels, dB) depends on many physical factors, but one of the largest effects is the distance between the AP and the device.

As a side effect of the probing process, we can estimate the location of each device based on the measured signal strengths between it and the nearby APs.

We have approximately 20 iPads in fixed locations throughout our office. Your goal is, given the actual location of a subset of the iPads as well as probing information, to estimate the location of the remaining iPads.

2 Data

You will be given 4 csv files. The data should be fairly clean and self-explanatory, but a complete schema will be included below. Note that the Signal Strength data contains AP-device pings over a 3 hour period. To simplify the data tables the timestamps have been removed, so there will be many "duplicate" entries which correspond to unique timestamps. Consider each entry to be unique and valid.

2.1 Tables

dc_ipads - This table contains the list of ipads for you to train on. Include identifiers for the iPads as well as their locations

dc_signal_strength - This table contains pings between the iPads and the Access Points as well as the signal strength of the ping. Remember, the apparent duplicate entries are actually unique pings at different times

dc_ap_locations - This table contains the list of access points as well as their locations.

dc_test_data - This table contains the list of iPads that you're to predict the locations of. The table also includes one ping for each of several Access Points for each iPad, so includes the relevant information about the ping and the Access Point.

2.2 Parameters

iPad Name - This is the name of the iPad device

iPad ID - This is a unique identifier for each iPad device

iPax x - X location of the iPad in meters in our building. (A corner of the building is set as (0,0))

iPax y - Y location of the iPad in meters in our building. (A corner of the building is set as (0,0))

Floor - Floor that the iPads are located on. (All iPads should be on the same floor: 3)

Access Point ID - This is a unique identifier for the Access Point

Signal Strength - For a given ping, the Signal Strength of the connection recorded by the Access Point

Access Point x - X location of the AP in meters in our building. (A corner of the building is set as (0,0))

Access Point y - Y location of the AP in meters in our building. (A corner of the building is set as (0,0))

3 Final Word

We understand that 2 hours is not long enough to get the best solution to this problem. We want to see your ability to come up with some initial ideas and demonstrate rapid prototyping. Keep in mind that this is just one

factor in your overall evaluation, but that your notebook is a great chance for you to demonstrate your thought process and problem-solving approach.

Good luck and do your best!

4 Problem Statement

Please return your predicted location of each ipad in the test data in a comma separated variable (.csv) file with the following column headings in the first line. A single row is given as an example here:

iPad ID, predicted x location in meters, predicted y location in meters
0452f3638d1a, 4.2, 32.5

Please send the notebook or code that you used to investigate the data and make the final prediction. Also include a brief description of your approach in the notebook.