E205 - Lab 1

Data Manipulation & Modeling

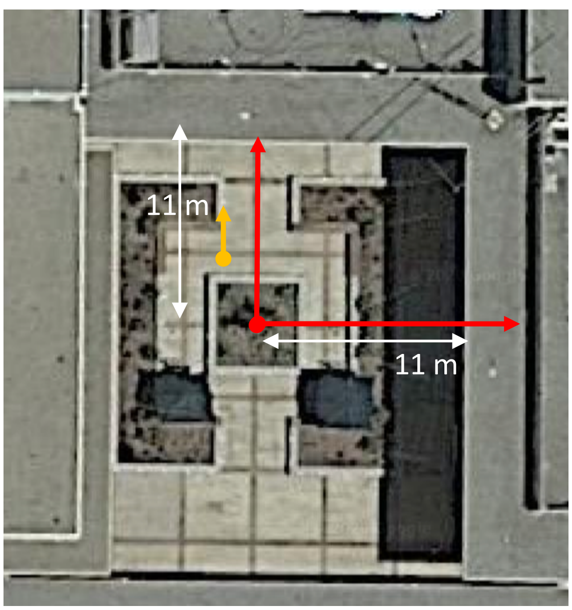
Due: 1:00pm Friday Sept. 11, 2020

1. Download the Data

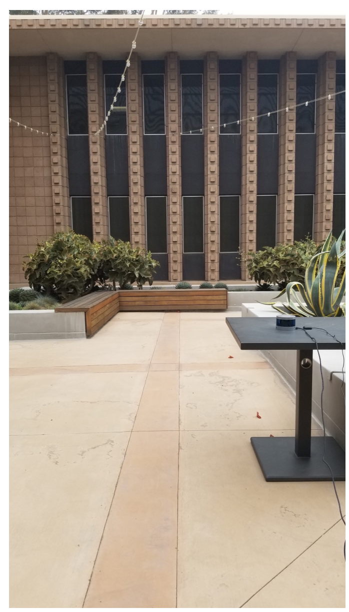
Start by downloading the lab1 csv files from the E205 web page. The data was collected with a Velodyne Puck VLP-16 3D laser scanner. The data in these files was obtained by setting the lidar in a stationary position in the Parsons building courtyard. The roll and pitch angles are approximately zero. The yaw angle is 90 degrees corresponding to facing north. See Figure 1.

Note the data files have columns of data in the following order: Range(m), Elevation(degrees), Azimuth(degrees), Range(m), Elevation(degrees), Azimuth(degrees), Reflectivity, Time Stamp(s) Latitude, Longitude, Time Stamp(s)

In the data sets, the only measurements logged are those with elevation angle zero, and one of three azimuth angles [-90, 0, +90].



(a)



(b)

Figure 1: Top down view (a) of Parsons building. Yellow arrow indicates (approximate) location and orientation of sensor, while the red axes represents an inertial Cartesian coordinate frame. In (b), an image of the lidar setup in the Parsons courtyard.

Also, download the additional data set police_killings.csv. There is a helper file written in python to load data into python if you like, called open_PKD.py. This data was downloaded from the website: https://www.kaggle.com/jpmiller/police-violence-in-the-us?select=police_killings.csv. We did not validate its accuracy or authenticity.

2. Create Histograms

Create and plot a histogram for each of the following data vectors:

a) range for azimuth = -90 degrees
b) range for azimuth = 0 degrees
c) range for azimuth = 90 degrees.

3. Create a Model

For the histogram in section 2 produced from the azimuth = -90 degrees data file, create a mathematical function that best characterizes the data and can be used as a conditional probability function P(Z|X). It should be a function of the actual distance to a wall or object. This mathematical function could be the sum of several other functions (e.g. a Gaussian function + a line), it could be piecewise, etc. There are several reasonable models given the data provided.

4. Transform and plot the GPS measurements

Transform the GPS longitude and latitude measurements (from the azimuth=0 file) to X and Y coordinates in meters. Recall we did this as part of E80. Define the origin to be located at the mean longitude and mean latitude measurements. Let the X-axis face directly east, and the Y-axis face directly north. Create a plot the 2D X, Y data set. Note the difference in variance in each direction.

5. Implement Bayes rule

Let us assume a very different coordinate frame that is centered in on the central planter in Parsons courtyard (see red coordinate frame in Fig. 1a). Assume the east and west walls are 22.0 meters distance apart in the east west direction, so that the origin of the coordinate frame is 11.0 meters distance to the nearest point on each wall. The closest point on the north wall is also 11.0 meters from the coordinate frame origin.

Now assume there are only four position states in which the sensor could be located. These states are located at:

$$\boldsymbol{x_1} = [x_1 \;\; y_1\,] = [+1.4 \;\; 2.8]$$
$$\boldsymbol{x_2} = [x_2 \;\; y_2\,] = [+1.3 \;\; 2.8]$$
$$\boldsymbol{x_3} = [x_3 \;\; y_3\,] = [+1.2 \;\; 2.8]$$
$$\boldsymbol{x_4} = [x_4 \;\; y_4\,] = [+1.1 \;\; 2.8]$$

If our prior knowledge of states is such that p($\boldsymbol{x_i}$) = 0.25, and we just obtained a range measurement that is the first -90 degree azimuth range measurement in the file (which we will call z), use Bayes rule to calculate the probability of being in position states 1, 2, 3, 4: i.e. p($\boldsymbol{x_1}$| z), p($\boldsymbol{x_2}$| z)), p($\boldsymbol{x_3}$| z), and p($\boldsymbol{x_4}$| z). Note: we are ignoring the other range data measurements for simplicity. Likewise, the y values are not significant here.

6. Create Police Data Model

Read the linked website for the police data, noting the various data sets available. Using the police_killings.csv data set, create a plot of the conditional probability functions:

p(victims_age | victims_race=black)
p(victims_age | victims_race=white)
p(victims_age | victims_race=hispanic)
p(victims_age | victims_race=asian)

These conditional probability functions can be described as the probability of a victim's age, given their race. The x-axis of the plot should be "age" in years. I suggest using 5 different age ranges (i.e. one could use 5 age ranges of 0-15, 16-20, 21-25, 25-30, 30+ years old). Be sure that the plots are normalized such that the area under the curve sums to 1.0.

7. Deliverables

A brief lab report should be submitted by 1:00pm Friday Feb 1 on Sakai under Lab 1 of *Assignments*. The report should include all histogram plots and mathematical plots listed above. A short few-sentence justification can be used for any modeling decisions. For problem 5, provide steps taken in calculating your solution.

Please use the style format of an IEEE conference paper format. You do not need to have the same sections as in most IEEE conf. papers, but you will for later labs.