# Hidden Markov Models Implementation

## Description –

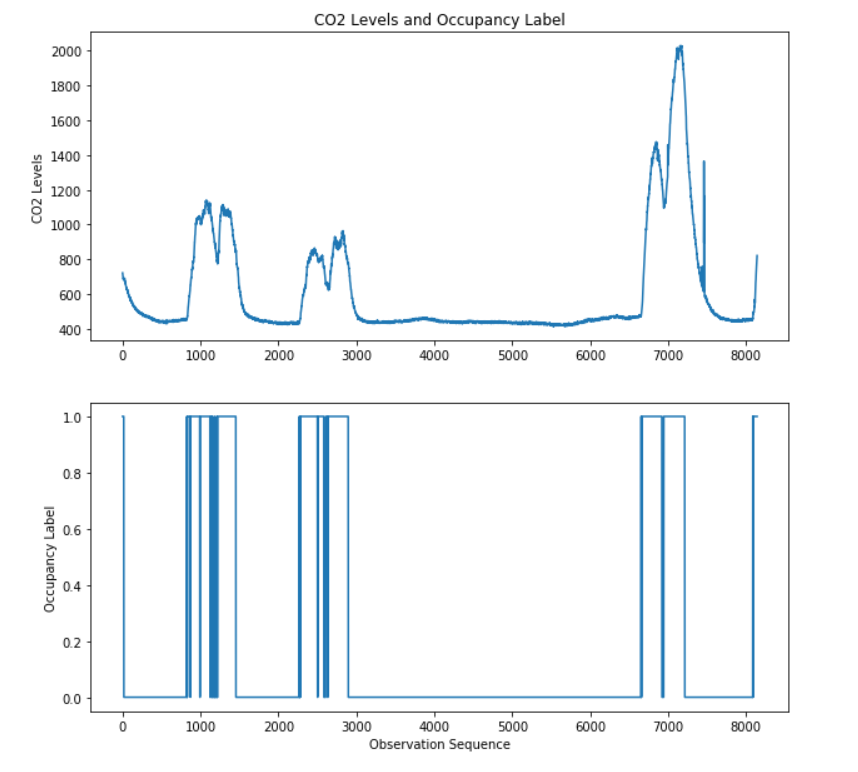
In this implementation, we are implementing Hidden Markov Model on Occupancy Detection Data Set. The data set contains features like CO2 level, Light, Temperature, Occupancy Status of the room. I am using CO2 levels and Light intensity one by one to predict the occupancy status of the room.

# Solution Method –

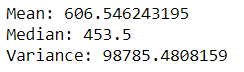
## Exploratory Analysis –

As mentioned in the project description I started with CO2 levels as my observation sequence and Occupancy label as the hidden states.

I plotted both the sequences with x axes depicting the observation sequence and y axes depicting the CO2 level and occupancy label. The figure below depicts the plot for the same.



After plotting both the CO2 levels and Occupancy labels I calculated the mean, median and variance for the CO2 levels. And the figure below shows the mean, median and variance value for the same.

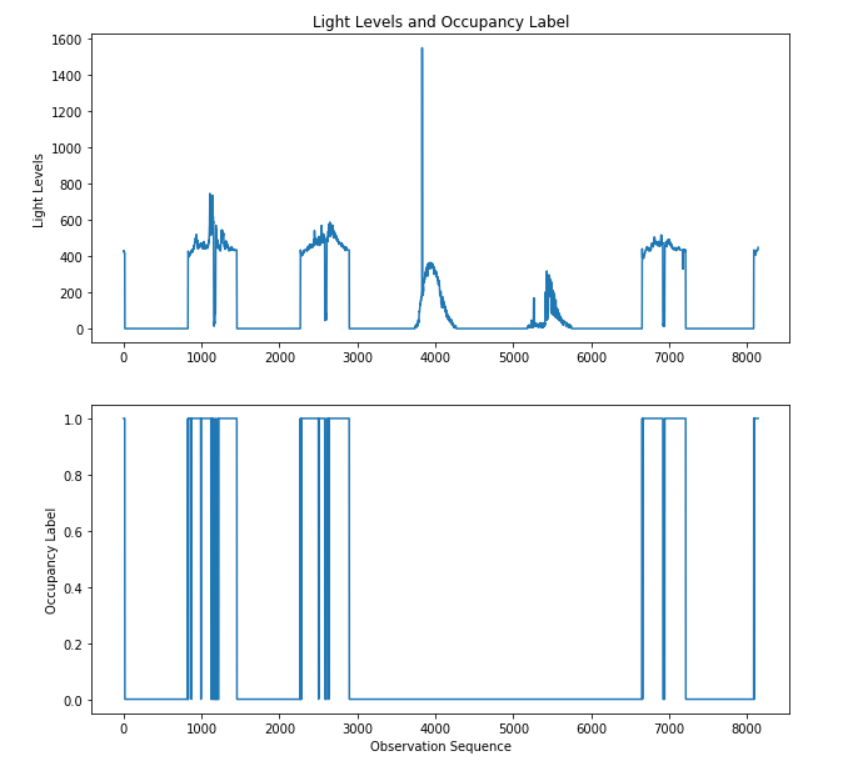


After that I calculated the dimension the dimension of the emission matrix. To do that, first the CO2 levels were needed to be classified into class labels otherwise there would be a huge number of unique observations and it would be hard for our Baum Welch algorithm to learn from that. So, I created 3 classes for the CO2 levels namely – Low, Medium and High depending on the levels of CO2. After this step I was left with 3 unique observations and 2 hidden states. So, the dimension of my emission matrix was 3 columns (for each unique observation) and 2 rows (for each hidden state).

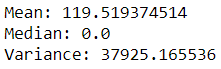
Since the data was recorded for almost every minute, I used the t of my observation as one minute. Initially, I had the concern with training the model with so many observations but I figured out that it wasn’t very hard to train with the entire observation sequence with 1-minute sampling. This sampling also gave me better results than the aggregated data.

The next task was to pick other variable from the remaining variable in the data set to predict the occupancy label. For that I used the value Light and the reason behind choosing Light as my other variable was, whenever the room is filled with people there is a high chance that lights would be switched on at that time.

After selecting my second variable for the prediction task, I plotted it with Occupancy label as an observation sequence. The figure below shows the same.



After plotting the chart, I calculate the mean, median and variance for the second variable which Light levels as well and the figure below reports the values.



After calculating these values. I again created the class labels for the Light level otherwise there would be a huge number of unique values of observations. So, I created 3 classes namely – low, medium and high depending on the value of light at a particular time. After this I selected the dimension of the emission matrix which is 3 columns (for each unique observation) and 2 rows (for each hidden state).

### Analysis –

For the analysis part, I implemented two Hidden Markov Models for each of the selected variable to predict the Occupancy label. I first used Baum Welch algorithm to train my models. I passed randomly generated values of starting probability, emission and transition matrix. After that I trained my observation sequences for 20 epochs each using Baum Welch algorithm.

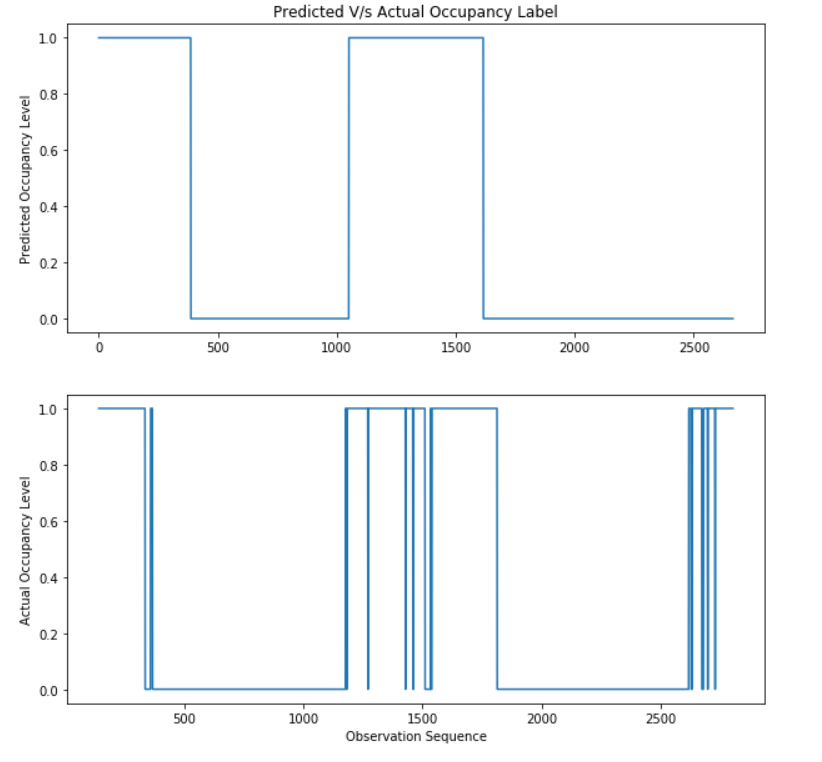
Once I trained my models I used Viterbi algorithm to predict the occupancy label for each of the trained HMM. The results for the prediction task is discussed in the next section.

## Result –

Both the Models – trained with CO2 level sequence, trained with Light level sequence were used to predict the Occupancy status. Both the algorithms gave very promising results. The results for each model are discussed below. For testing the models were tested on test data set.

### Model 1 – With CO2 Levels –

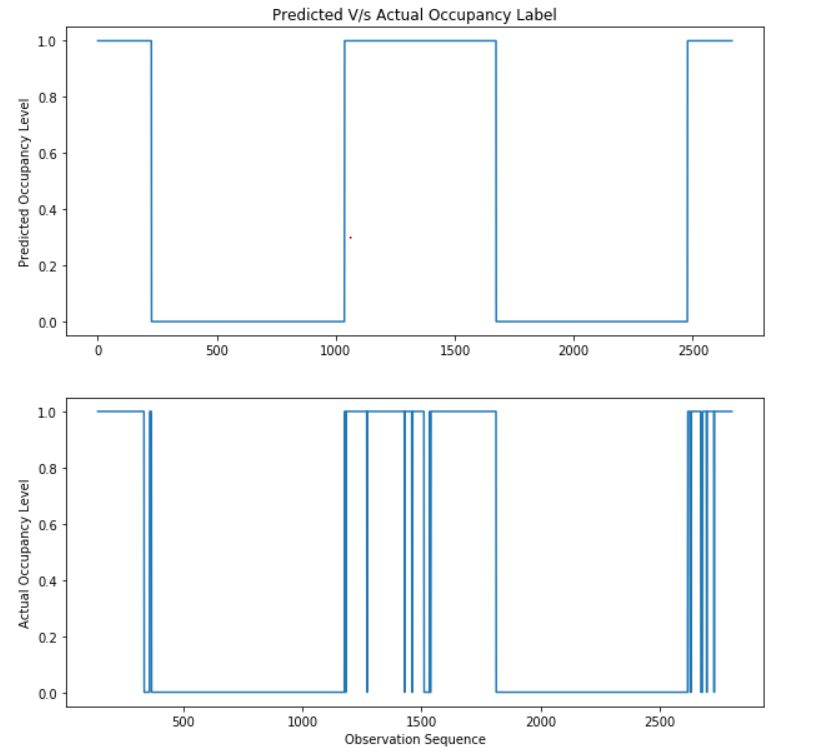
Accuracy with this model is around **82.90%.** The model performs reasonably. The figure below shows the predicted vs the actual sequence of Occupancy level on the train set.



### Model 1 – With CO2 Levels –

Accuracy with second model trained with the sequence of Light levels gives a very high accuracy of **97.03%.** The performs excellent using the level of Light sequence.

The figure below shows the predicted vs actual value Occupancy label predicted using model 2.



It can be seen that the model performs excellently and gives amazing result with model 2.