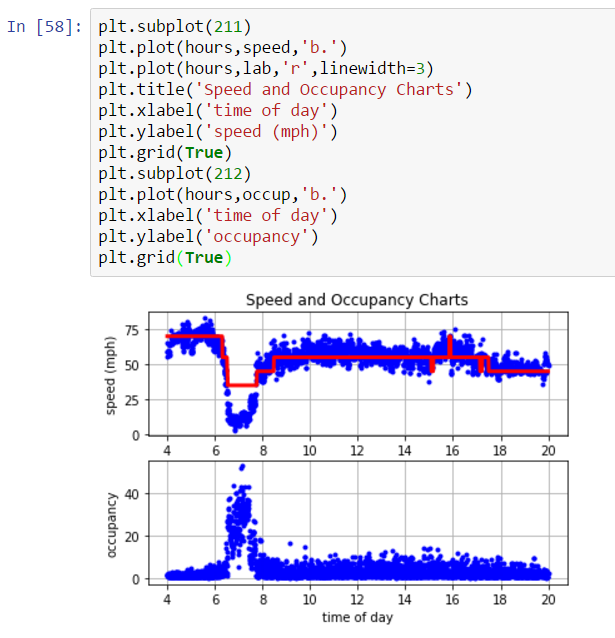
# CIVE-650C –Homework Data Smoothing, Classification, and Prediction

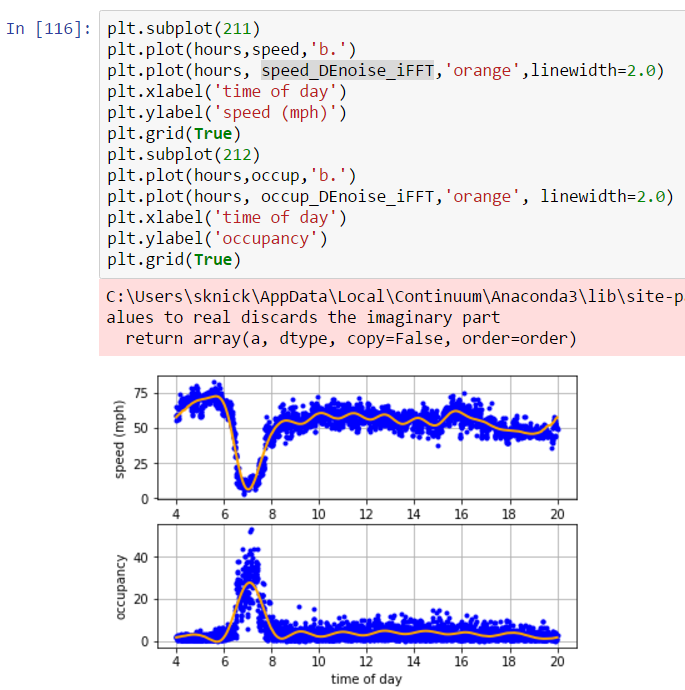
Skylar Knickerbocker

Task 1 — Plot the raw data (VSL\_Data1.csv), with proper legends, title, axis, etc.



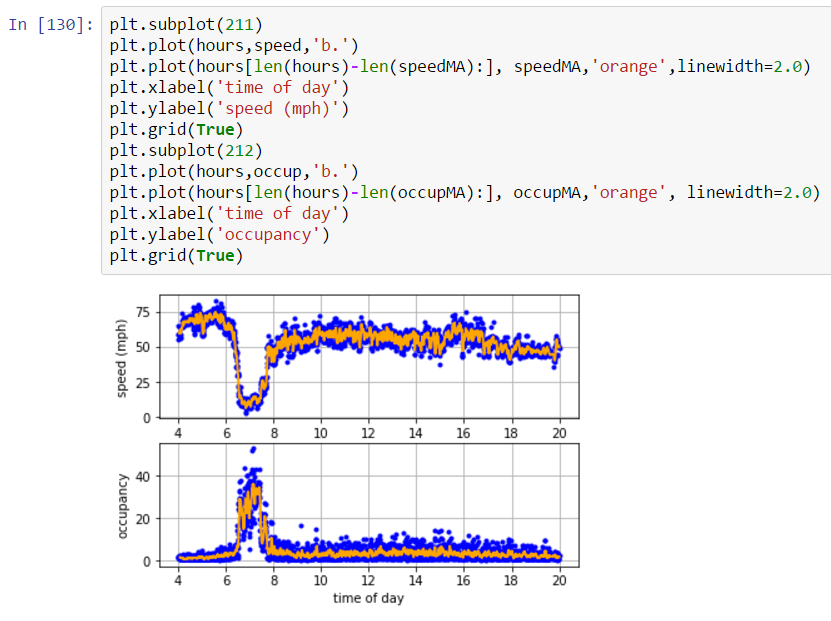
Task 2 — Data Smoothing

1) FFT Smoothing As discussed during the class, plot the smoothed speed and occupancy.



2) Another Choose another data smoothing method, plot the smoothed speed and occupancy.

**Used a simple moving average of previous 6 readings (2 minutes)**



Task 3 — Classification

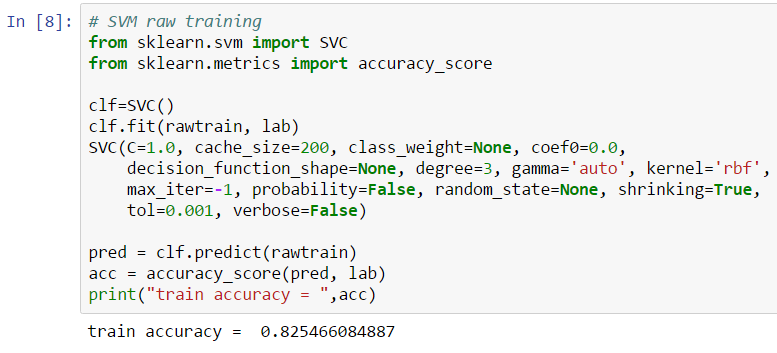
Choose TWO different classification algorithms (A, B) and train the model, Label = f(speed, occu), using 3 different dataset; therefore one should have 6 different models:

**Chose to use the SVM and Decision Tree classification algorithms**

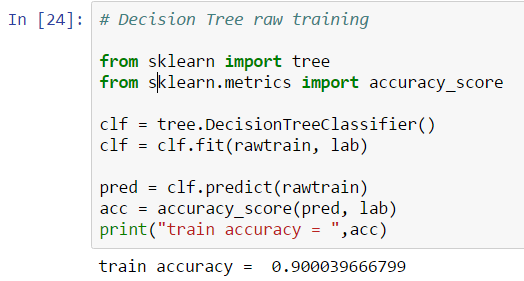
**Summary of accuracy findings**

|  |  |
| --- | --- |
| **Classification** | **Accuracy** |
| **SVM using Raw Data** | 0.825466084887 |
| **Decision Tree using Raw Data** | 0.900039666799 |
| **SVM using FFT** | 0.966283220944 |
| **Decision Tree using FFT** | 1.0 |
| **SVM using Moving Average(MA)** | 0.880365659777 |
| **Decision Tree using Moving Average (MA)** | 0.99960254372 |

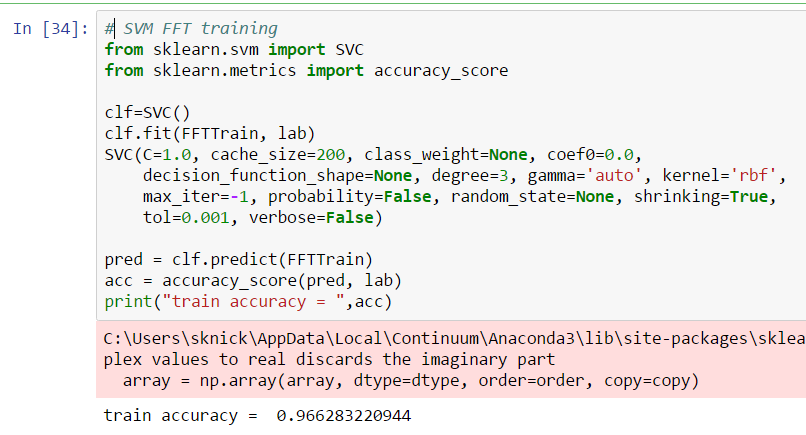
1. **SVM using Raw data**

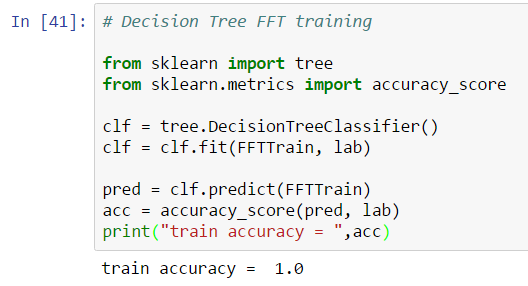


1. **Decision Tree using Raw data**

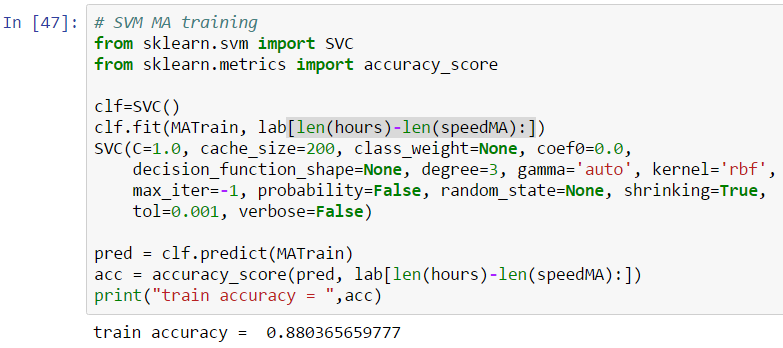


1. **SVM using FFT**

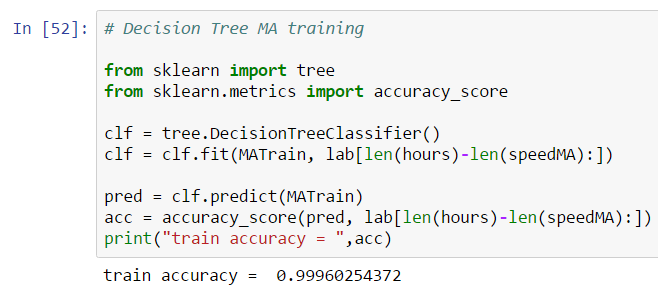


1. **Decision Tree using FFT**

1. **SVM using MA**



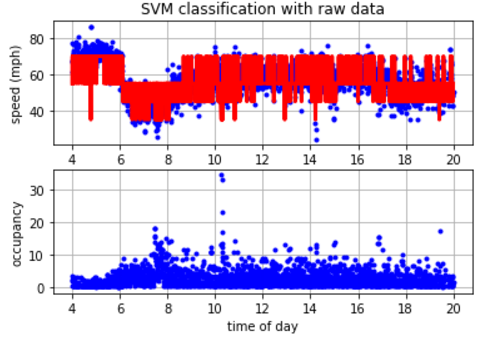
1. **Decision Tree using MA**



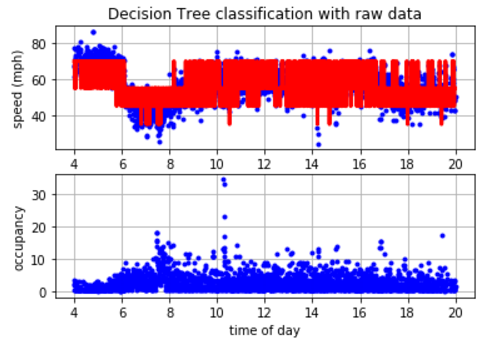
Task 4 — Prediction

Once you have all 6 models ready, apply them on the new data set VSL\_Data2.csv

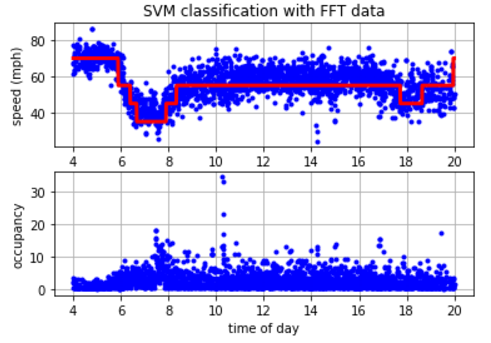
1. **Prediction using SVM classification with raw data**



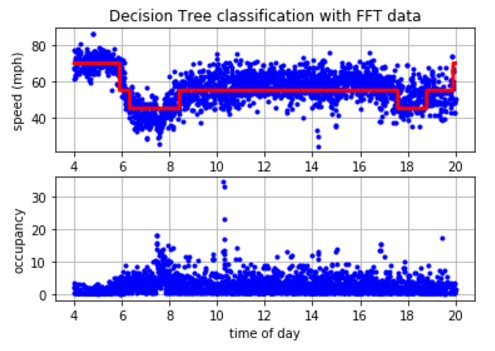
1. **Prediction using Decision Tree classification with raw data**



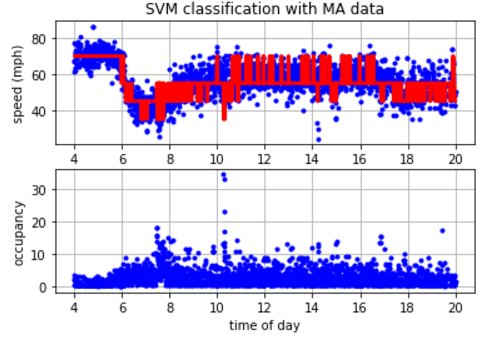
1. **Prediction using SVM classification with FFT**



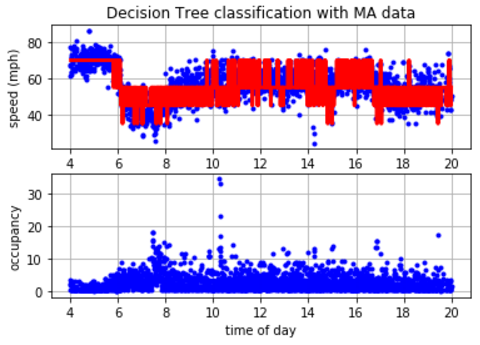
1. **Prediction using Decision Tree classification with FFT**



1. **Prediction using SVM classification with MA**



1. **Prediction using Decision Tree classification with MA**



**Discussion:**

Out of the six models, the models using FFT smoothing performed significantly better. The training using raw data resulted in high fluctuation in the classification. The Decision Tree had a higher accuracy after training but after applying the prediction to the test dataset showed even greater fluctuation in the classification resulting in an unusable output. The moving average smoothing had very little improvement in the classification compared to raw. Both had higher accuracy rates after training with the Decision Tree classification getting 99.9%. Unfortunately both models had high fluctuation in the classification with the SVM classification performing slightly better. Overall the best classification was achieved using the FFT smoothing. Both models had high accuracy after training including 100% with the Decision Tree. Both classification predictions showed excellent results. Both classifcations captured the dip in speeds during the AM and PM as well as throughout the day. The SVM model performed slightly better in the ability to classify lower speeds compared to the Decision Tree in the AM peak. Both classification had approximetly the same activation during in the AM and PM with the Decision Tree lasting slightly longer in the PM. Either model would be acceptable but the SVM using FFT would ultimately be chosen based on the high accuracy and ability to classify events correctly.