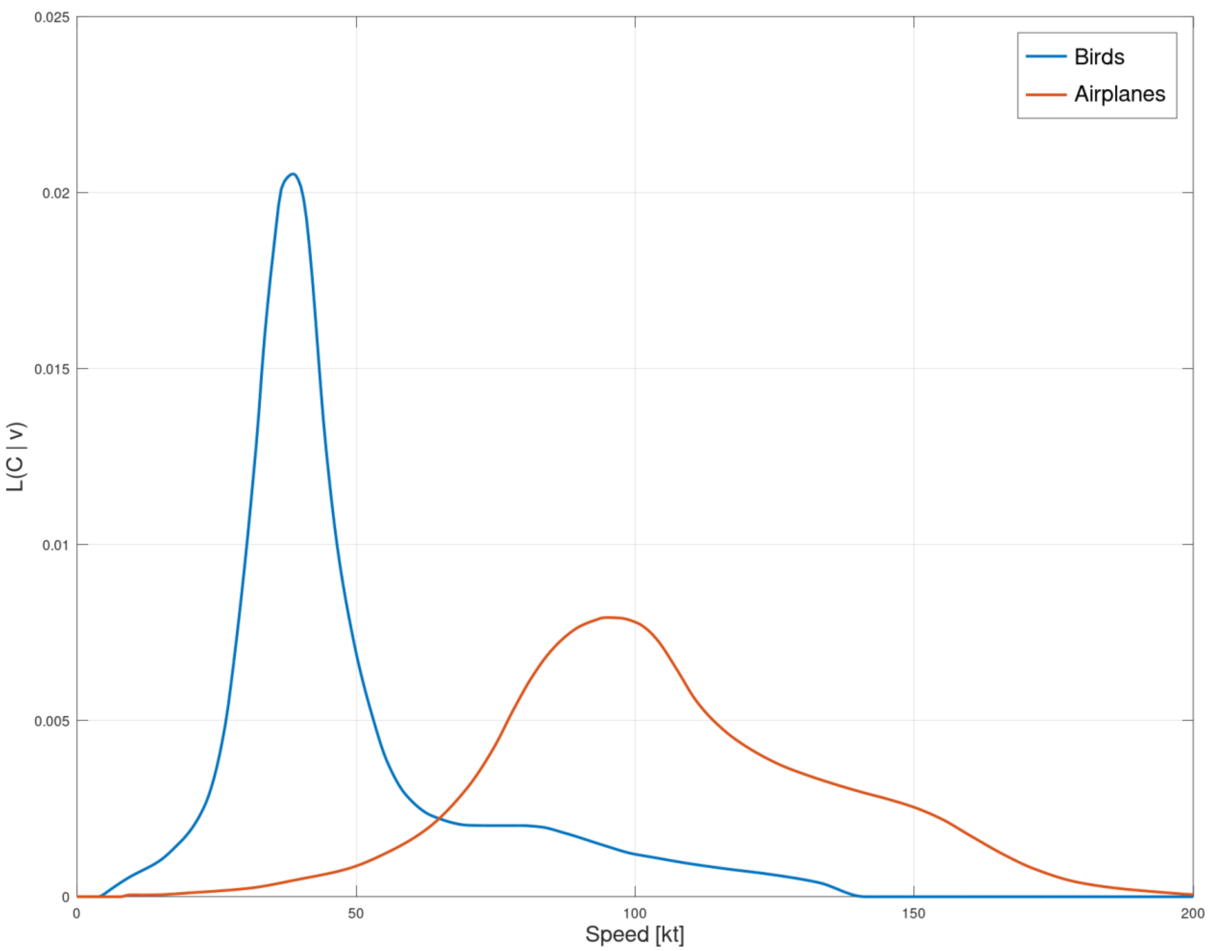
Naïve Bayesian Classification

A RADAR TRACE CLASSIFIER (100 Points)

A frequent problem at airports is the collision between airplanes and birds. You are to solve this problem by classifying radar tracks into two classes: birds and airplanes. Using a Naïve Recursive Bayesian classifier, your job is to calculate and report the probability that the object belongs to one of the two classes for **each data point provided**.

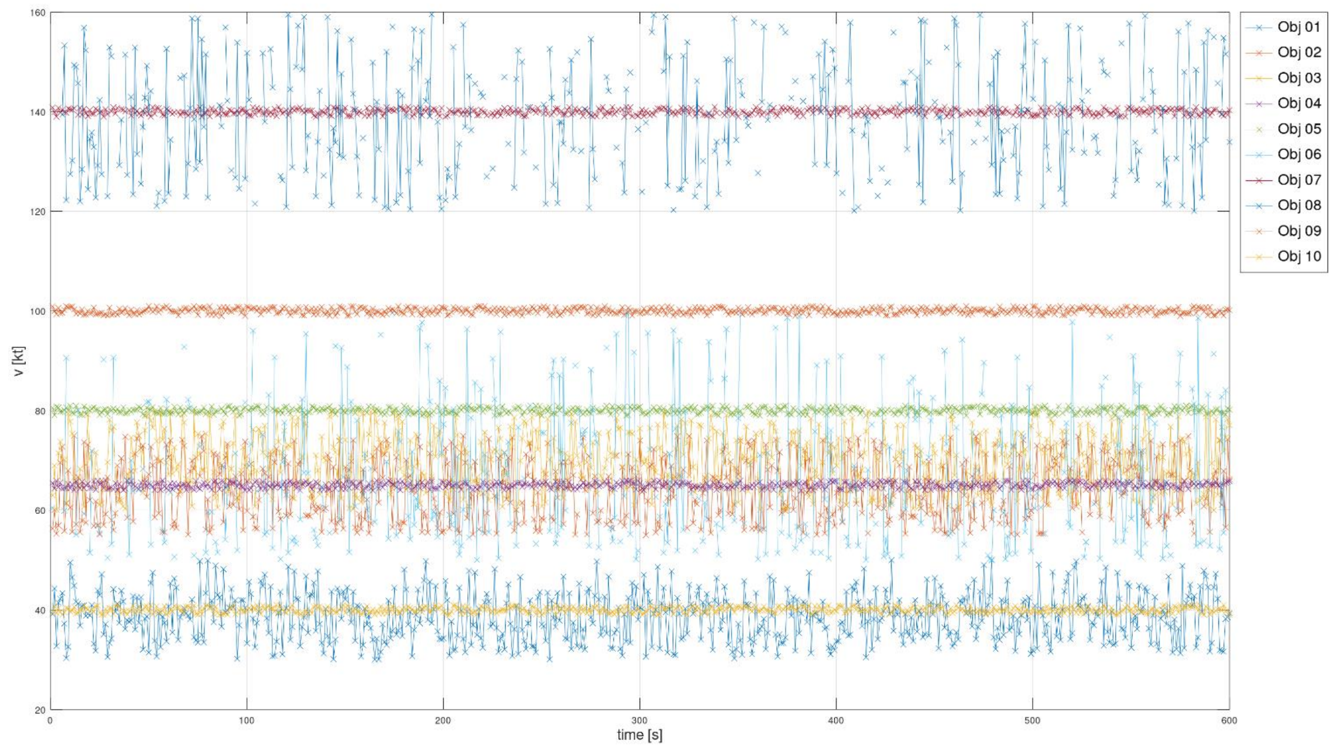
For your classification, you are given the following data:

a) The likelihood (`likelihood.txt` file) of **birds** (first row in the file) and **airplanes** (second row in the file) for specified speeds is represented here:



b) Twenty (20) tracks (`dataset.txt` file) representing the velocity of the **birds** and **airplanes** (10 rows of birds followed by 10 rows of airplanes in the training file) measured by a military-grade radar (1s sampling frequency for a total length of 600s). The corresponding data point would be a **NaN** value if the radar could not acquire the target and perform the measurement. These tracks are curated to have a maximum sample drop rate of 5% of the total number of samples per track.

c) Ten (10) tracks (`testing.txt` file) represent the velocity of the unidentified flying object measured by a military-grade radar (1s sampling frequency for a total length of 600s). If the radar could not acquire the target and perform the measurement, the corresponding data point would be a NaN value. These tracks are raw data.



For each test track, your solution will have to return a classification ("a" for **airplane**, "b" for **bird**) for **each data sample** and a **final classification** that summarizes the object class for the entire track.

Assume that the classifier is conservative when transitioning between classes of objects. A transition probability $P(C_{t+1} = bird \mid C_t = bird) = 0.9$ and $P(C_{t+1} = airplane \mid C_t = airplane) = 0.9$ should be sufficient. However, feel free to change these values as appropriate. These transition probabilities should remain fixed throughout the execution of your solution.

As initial probabilities for the classes, it is normal practice to start the classification from equal priors (for two classes, it would be 0.5 for each class). Expect these values to be constant as they are applied only to the first cycle of your estimation loop. Conversely, the posterior probability will change over time as the classifier acquires more signal information.

A solution considering the target's velocity as the sole classification feature is only relatively accurate by design. In fact, while testing your application, you will notice that, in certain cases, the object's speed alone is insufficient to make a reasonable determination as data is insufficient or too noisy. You must design an additional feature observable in the testing data to improve the classification. Use the traces in the data provided in the `dataset.txt` file to extract this new feature and produce its likelihood distribution for **airplane** and **bird** classes. Make sure to explain your rationale in the README file. For the assignment to be complete, you must modify your original classifier to include this new feature and the corresponding likelihood.

SUBMISSION

Your solution may use external libraries for pre-processing, fundamental calculations (e.g., linear algebra), and visualization. However, you must implement the core portion of your solution from scratch.

Python or C++ are the preferred implementation languages. If you are writing in C++, along with your source code, please include the main CMake file as well as any other instructions needed to compile the executable. For Python, provide only plain PY files (no Jupyter Notebook or PYC cache files).

Submit a single ZIP file via Canvas that includes all the source files and a README clearly explaining your solution's assumptions.

SOLUTION

$O_1 = \boldsymbol{bird}, O_2 = \boldsymbol{bird}, O_3 = \boldsymbol{bird}, O_4 = \boldsymbol{airplane}, O_5 = \boldsymbol{airplane}, O_6 = \boldsymbol{bird},$
$O_7 = \boldsymbol{airplane}, O_8 = \boldsymbol{airplane}, O_9 = \boldsymbol{airplane}, O_{10} = \boldsymbol{bird}$