This folder contains all the code for analysis and the implementation of the motif discovery algorithm BASS.

For questions, email *gautam\_nallamalaATfas.harvard.edu*

**Behavioral Action Sequence Segmentation (BASS) usage:**

Three files are used for BASS:

bass.pyx

main\_synthetic\_dataset.py

setup\_bass.py

To run the algorithm, Python3, Cython, NumPy, SciPy are required. editdistance is also required, which can be installed using the command:

*pip3 install editdistance*

Before running any of the code, bass.pyx has to be compiled using:

*python3 setup\_bass.py build\_ext --inplace*

To run a test case, use:

*python3 main\_synthetic\_data.py 7 50 5 10000 0.0 0.0 0.15 4 0.0 1 &*

bass.pyx contains the implementation of the motif discovery algorithm, the specification of the mixture model and miscellaneous functions used for the analysis in the paper.

main\_synthetic\_dataset.py is a sample application of the algorithm. This code generates a synthetic dictionary of motifs, a dataset from that dictionary and applies the motif discovery algorithm on the dataset. The output files contain the true dictionary and the one learned by the algorithm along with the probabilities of each motif.

**Important:**

For each new application, a `soft’ clustering model has to be specified. In the code, this is implemented in the GMM\_synthetic class in bass.pyx. This class has to be appropriately modified or alternatively a new class should be defined which contains the two functions defined for this class – “\_compute\_likelihood” and “\_generate\_sample\_from\_state”.

**Analysis of tracked zebrafish larvae data:**

Raw data is in the four folders

resultsMay2019

Catamaran\_pH\_2bTxtOnly

Catamaran\_pH\_2cTxtOnly

resultsForSB1

The code used for analysis of the tracked fish data is included in the Jupyter notebook titled “Zebrafish\_larvae\_analysis\_acid\_data\_final.ipynb”. This includes the extraction of the tracking data, computing the various parameters for each bout – speed, delta heading, duration of bout, tail amplitude, etc. Also includes clustering and generation of various plots presented in the paper.

The three files named acid\_\*.npy contain the mixture model parameters presented in the paper.

The files named “data\_explo\_hmm.npy”, “data\_dbsharppH\_hmm.npy”, “lengths\_explo\_hmm.npy” and “lengths\_dbsharppH\_hmm.npy” are reduced versions of the dataset containing only the information used for clustering and BASS.

**Using BASS to reproduce results from pH experiments:**

The Python code to run BASS on the exploration and aversive environment data is in

main\_explo.py

main\_acid.py

These take input the Gaussian Mixture Model learned in the previous section, whose parameters are stored in the three files named acid\_\*.npy. They also require the four data files “data\_explo\_hmm.npy”, “data\_dbsharppH\_hmm.npy”, “lengths\_explo\_hmm.npy” and “lengths\_dbsharppH\_hmm.npy” generated from the analysis.

The parameters used in the paper are eps = 0.1, p\_d = 0.2, Jthr = 0.15. To start with, we recommend running these two files with parameter eps = 0 since it is quicker to converge. The rest of the parameters are unused except for w\_thr (anything less than 1e-3 works equally well, setting too low takes very long time to converge).

The output from BASS can be read by another Jupyter notebook, see below.

**Analysis of output from BASS:**

The code used to analyze the output from BASS is included in “analyze BASS output files, final version.ipynb”. This includes extracting the dictionary of motifs in exploratory and aversive environments and generating the various plots presented in the paper.

Some of the files generated in this code are used as input for the “Zebrafish\_larvae\_analysis\_acid\_data\_final.ipynb” notebook (see comments).

The “motifs\_compare” folder contains the dictionaries obtained by running BASS on the exploratory and aversive datasets (five trials each).