

# Unsupervised behavior estimation

Luis Nicolas Luarte Rodriguez

## State of the art

The successful mapping of neuronal activity to behavior is one of the main goals of neuroscience. Behavior can be understood at three levels (1) computationally, that is, ‘why’ a specific behavior came up in evolution for a particular species; (2) algorithmically, ‘what’ the behavior does or what is the underlying algorithm; (3) implementation, ‘how’ such behavior is implemented at a neural level (Krakauer et al. 2017). A comprehensive view of all three levels is necessary for neuroscience to generate and test a reasonable hypothesis. However, while there are multiple analytical tools for each one, levels 1 and 2 have taken the back seat in this regard. One of the main requirements of those levels, which correspond to the more ‘observable’ ones, is quantifying animal behavior. In rodents, one of the leading experimental models in neuroscience (Ellenbroek and Youn 2016), obtaining behavioral data is usually done manually, tedious, slow, and prone to error and bias (Huang et al. 2020).

The main alternatives to manual labeling are supervised and unsupervised statistical approaches. The supervised method relies on previously annotated data, while the unsupervised does not. To achieve this, unsupervised methods leverage on algorithms that can extract statistical features from data and deduce different behavioral patterns (Gris, Coutu, and Gris 2017). This process has been made to identify rodent behavioral motifs across multiple time-scales (Shan and Mason 2020). More complex implementations of Hidden Markov Models using 3D pose estimation exist and have adequately described behavioral ‘modules’ or ‘syllables’ that are composed into more complex behavior (Wiltschko et al. 2015). Another approach to classifying rodent behavior is to use morphological operations and image segmentation to obtain high-resolution rodent ‘shapes,’ which can classify social interaction, for example, when two shapes collide (Unger et al. 2017).

Previous unsupervised approaches are heavily reliant on image processing. While video recording of animal behavior does not represent a considerable cost, and image processing is fast in modern hardware, it may not always be the case that obtaining video footage is possible. To address these kinds of situations, relying on a few measurement devices such as accelerometers might prove useful as it can be decomposed into multiple parameters such as acceleration, static

acceleration, dynamic, among others. These features can be clustered using expectation maximization fitting of Gaussian Mixture Models(Chimienti et al. 2016), resulting in the identification of distinct behaviors that can be later on labeled, based on correlation with previously associated feature-behavior pairing or by expertise.

Generating workflows that speed up behavior quantification can augment such data's resolution, making it possible to pair it with neuro-related data acquisition such as electroencephalogram and functional magnetic resonance. Neuroscientists have readily identified this, and tools are being developed to ease the annotation process (Pereira, Shaevitz, and Murthy 2020). Unsupervised techniques seem very promising as they require almost no human intervention, can work with a small number of features, and most algorithms are not computationally expensive so that they can be implemented on standard hardware.

## Hypothesis

This final project explores options for manual annotation of rat behavioral data by using unsupervised clustering techniques. Video footage of rat behavior was obtained from Gerós, Magalhães, and Aguiar (2020) dataset, which is composed of 10~ minutes sequences of rats in an open-field arena without any intervention. Four fully annotated (with seven behaviors) sequences were extracted and processed with the use of custom software to obtain a position (x, y coordinates), on a frame by frame basis, of the head, body ('hips'), and the tip of the tail. The central hypothesis is that simple feature construction from position coordinate and clustering techniques can achieve acceptable classification accuracy levels compared to manually annotated data. Acceptable accuracy levels will be considered true positives plus true negatives overall classification combinations over 60%, obtained via cross-validation procedure. All data processing from rat position extraction to classification is done entirely unsupervised. The supervised part is only for validation of the proposed technique.

## Objectives

### General

1. Measure the performance of an unsupervised approach to behavior classification against manually annotated data on a frame by frame basis.
2. Use clustering algorithms to classify distinct behaviors and correlate them with the manually annotated labels.
3. Fit a classifier with clusters and handcrafted features to predict actual labels and measure performance.

## Specific

1. Use k-means, and Gaussian Mixture Models to cluster position and distances traveled data to obtain unsupervised labels.
2. Correlate unsupervised labels with manually annotated labels using Chi-square test and Crammer's V test over binned data.
3. Fit a K nearest neighbors classifier, using clusters and handcrafted features for every rat to obtain accuracy and derive unsupervised technique performance.

## Methods

The first procedure was to obtain clusters for every frame, based on distance, rolling average of Speed, and rat position. The first algorithm was applied as K-means. This algorithm aims to generate an optimal data partition into 'k' clusters so that points within a cluster are close together, whereas points from other clusters are far apart. K-means is a problem where the goal is to minimize the total intra-cluster variance:

$$\text{Function to minimize} = \sum_{j=1}^k \sum_{i=1}^n \|x_i - c_j\|$$

Where  $k$  is the number of clusters,  $c$  indicates the cluster,  $x$  the individual data point, and  $n$  are the cases, the algorithm proceeds by selecting  $k$  random cluster centers, then it assigns each case to the closest center, considering the above euclidean distance. It calculates the mean and re-assigns all the points to the closes mean or centroid for every cluster. Finally, it calculates the objective function and minimizes it over multiple random iterations, so the iterations with the lowest total within-cluster variance are picked, and its centroids are used for clustering. K-means was computed for every rat; qualitatively, centroids grouped data in the 'center' and 'corners' (fig. 6), which is significant as behaviors in open-field test tend to group around the center and corners, more locomotor activity within the center signals explorations. In contrast, activity near the corners points to anxiety (Sestakova et al. 2013). It should be noted that there are extensions for k-means for Spatio-temporal data (Ossama, Mokhtar, and El-Sharkawi 2011). However, this is not implemented in popular programming languages.

Finally, the number of centers to fit was determined for each rat using the 'Silhouette Score,' which considers the (a) the mean distance between a given point and all other points belonging to the same cluster and (b) the mean distance between the observation and all other points from the next closes cluster:

$$SS = \frac{(b - a)}{\max(a, b)}$$

Where SS represents the 'Silhouette Score' ranging from -1 to 1, being one a

dense and precise segmentation, whereas 0 represents overlapping clusters and -1 possibly indicates that points are not well assigned to clusters.

Motivated by the bimodal distribution that some of the features present, Gaussian Mixture Models were chosen as candidates for clustering. A mixture model is basically defined as a distribution composed by multiple distributions and a ‘mixing’ factor:

$$G(x) = \sum_{n=1}^N \alpha G_k(x)$$

Where  $G$  represents the mixture distribution,  $\alpha$  the mixture factor, and  $G_k(x)$  are the distributions, in this case, the will be gaussian, each having its parameters. Similar to k-means, mixture models require the number of components to be specified. The R programming language offers the ‘mclust’ package that uses the Bayesian Information Criterion to find the optimal number of clusters while preventing overfitting. Similarly, initialization points are determined by a hierarchical clustering algorithm. Unlike k-means results, visualization is less clear on clustering space (fig. ??).

The next step was to compare the two sets of clusters given by k-means and Gaussian Mixture Models when clusters were obtained. The first approach was to compute the correlation between the clusters and the annotated labels using the chi-square test, and Cramer’s V. Cramer’s V is typically used to assess association strength after the chi-square test has proven significance. This way, it could be determined is the correlation exists and to what extent.

Gaussian Mixture Models were chosen (see results section) as clusters. The main idea was to fit a model considering clusters as an interaction term, so the classification depended on the cluster. For the independent variable ‘rolling speed’ (see data description) was selected, no further terms were added as Speed included information about all other terms. As for the model, K Nearest Neighbors was selected (1) because of its low computational requirements, which could prove useful for live video processing; (2) non-parametric, because the problem of predicting labels is stated without knowledge of data distribution and (3) its interpretation is relatively straightforward. This unsupervised algorithm assumes that similar things are closer together in feature space. Finally, given that the expected data size is small, K Nearest Neighbors should not run into performance issues regarding classification speed. The algorithm was applied to each rat separately as data is not independent, and the goal of the method is to generate predictions in a live setting, that is, without having extensive datasets. Nevertheless, a global ‘model’ was applied to all data for comparison. Algorithm validation was through repeated cross validation, considering accuracy as a primary performance metric.

# Results

## Clusters

The number of clusters was determined for each rat independently but typically ranged from 6 to 9. As for the cluster distribution over rats, there were no clear trends (see fig. 8 and fig. 9, for k means and GMM, respectively).

## Correlations

For all contingency tables, of both k-means and GMM clusters, p-values < 0.001. Regarding the association's strength, Cramer's V showed a slight advantage for GMM, especially in rat 2 and 4 (see more details in the table below).

| Rat Number | K-Means | GMM  |
|------------|---------|------|
| 1          | 0.32    | 0.25 |
| 2          | 0.34    | 0.46 |
| 3          | 0.18    | 0.18 |
| 4          | 0.25    | 0.39 |

## Classification performance

As previously stated, KNN was chosen as the classification algorithm; results reported below were obtained using repeated cross validation (3 repeats), data was centered and scaled.

| Rat Number | Accuracy | Kappa | K hyper-parameter |
|------------|----------|-------|-------------------|
| 1          | 0.64     | 0.25  | 5                 |
| 2          | 0.81     | 0.62  | 15                |
| 3          | 0.57     | 0.14  | 19                |
| 4          | 0.81     | 0.62  | 7                 |

As expected, accuracy was higher in rats that reached higher correlation strength. While accuracies are moderate to good, the Kappa statistic, which indicates agreement between the classifier and data beyond chance, only ranges from slight to moderate. This could be explained upon further inspection of the confusion matrix.

|   | 0    | 1    |
|---|------|------|
| 0 | 15.7 | 28.6 |
| 1 | 7.2  | 48.4 |

Class prediction imbalances are notorious for class ‘0,’ that is, stationery labels, which makes sense given that non-stationary behaviors are more involved in terms of the paw and head movement, features that are not considered by the extraction software.

## Data description

Raw data consist of frames recording rat in infrared light (fig. 1). These data are processed with custom software to obtain the position of the rat’s head, body, and tail. Processed data consists of the rat’s path during the open field task (fig. 2). One of the main handcrafted features is the distance traveled, which is calculated as the euclidean distance between point with lag 1, that is, with the previous frame (fig. 3). With distance computed, Speed was calculated considering the reported 30 frames per second (FPS) (Gerós, Magalhães, and Aguiar 2020), that is,  $distance/0.5sec$ , thus reported in arbitrary units per second. As time changes are constant, distance and Speed are almost equivalent, so the rolling average over 1 second (30 FPS) was considered (fig. 4).

Regarding distances, all rats show two apparent density ‘peaks’ around short and large distances. These densities correlate with the behavioral annotations, which can be divided into movement-related and stationary. In fact, for all analysis annotations, we re-coded from 7 behavior to two, representing movement-related and stationary (see table below). Re-coded labels showed similar distributions in all rats (fig. 5).

| Coded | Behavior            | Re-coding        |
|-------|---------------------|------------------|
| 1     | Resting             | Stationary       |
| 2     | Walking             | Movement-related |
| 3     | Moving exploration  | Movement-related |
| 4     | Local exploration   | Movement-related |
| 5     | Supported rearing   | Stationary       |
| 6     | Unsupported rearing | Stationary       |
| 7     | Grooming            | Stationary       |

## Data pre-processing

For all procedures, features were scaled, and missing values were dropped or ignored. For K nearest neighbors fitting data was down-scaled to account for slight class imbalances. As the dataset can be considered repeated measures, all analyses were performed on individual rats, and rat stratified cross-validation procedures.

## Conclusions

The main objective was to explore techniques that could ease the burden of manual annotation of rodent behavior. The data processing pipeline started with (1) automatic extraction of features from video recording (not shown here, done by custom software); (2) handcrafting simple features derived from the previous ones, such as Speed and distance traveled; (3) use all the features to generate clusters, the intuition here is that clusters should represent areas where rat performed more actions, and in that regard better method exists such a previously mention extensions to k-means for Spatio-temporal data; (4) fit a model or classifier to each of the clusters adding an important feature, such as Speed, here the main idea is that movement could reflect different behavioral patterns depending on the place (corners, edges, center, etc.). Finally, (5) using the fitted model or classifier could help in labeling unannotated data.

The main limitation of this initial approach was the small number of rats, the clustering algorithms were not explicitly designed for Spatio-temporal data, and the results were mixed. A more sophisticated statistical method could be employed in the same workflow to improve results. Interesting additions could be applying fast convolutions in the image processing stage to obtain more features related to torsion or body angle.

## Appendix



Figure 1: Raw data consists of multiple frame recording mice movement on infrared lighting

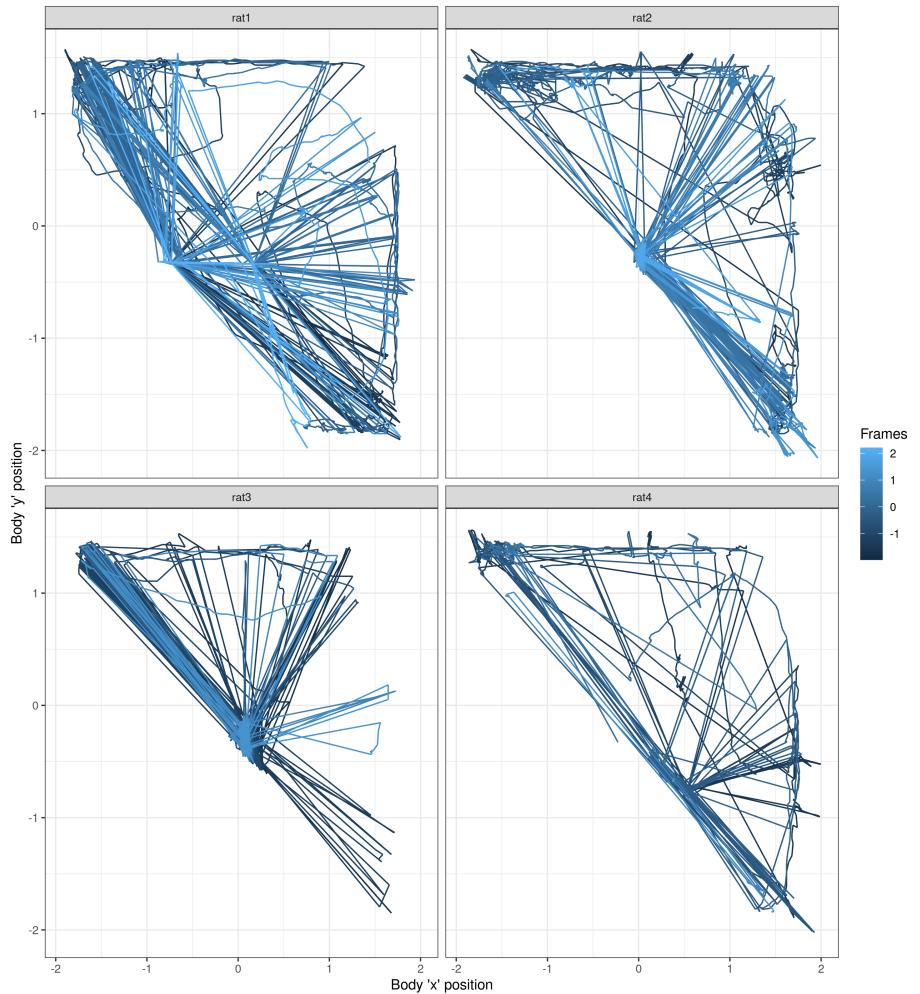


Figure 2: Path data from all rats for all recorded frames

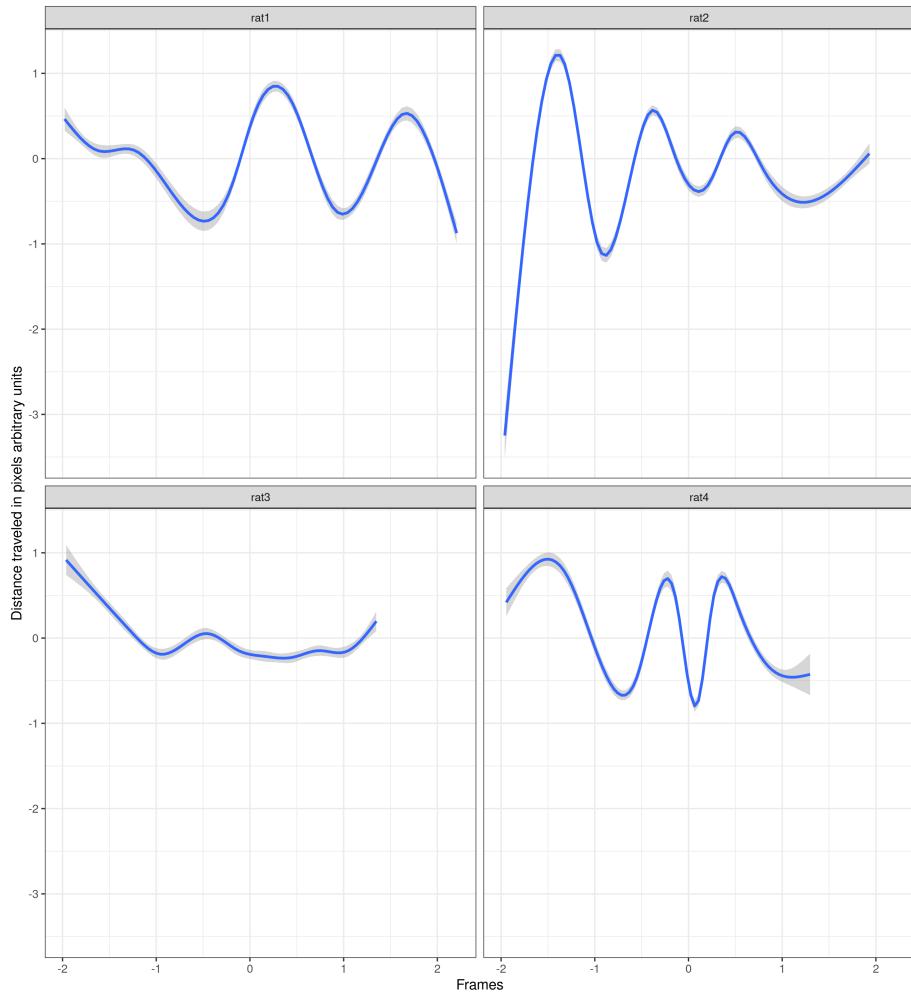


Figure 3: Distance traveled in pixels arbitrary units over frames

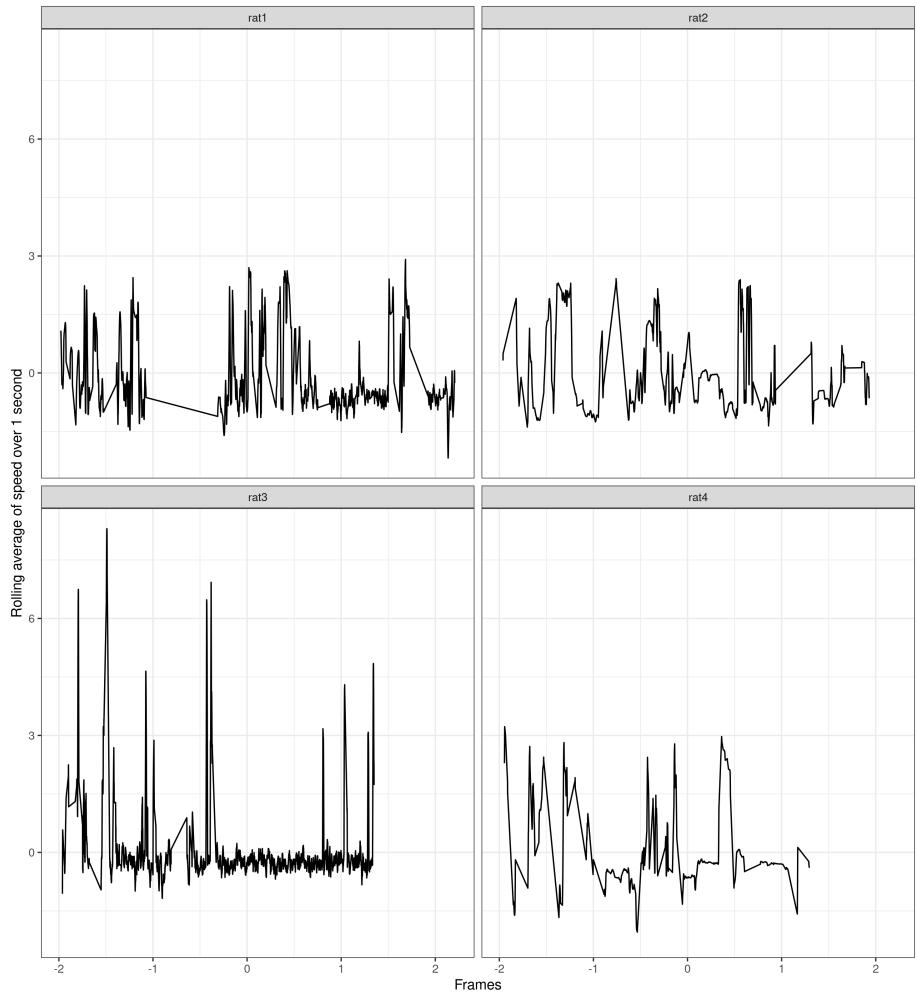


Figure 4: Speed was considered as a rolling average over 1-second

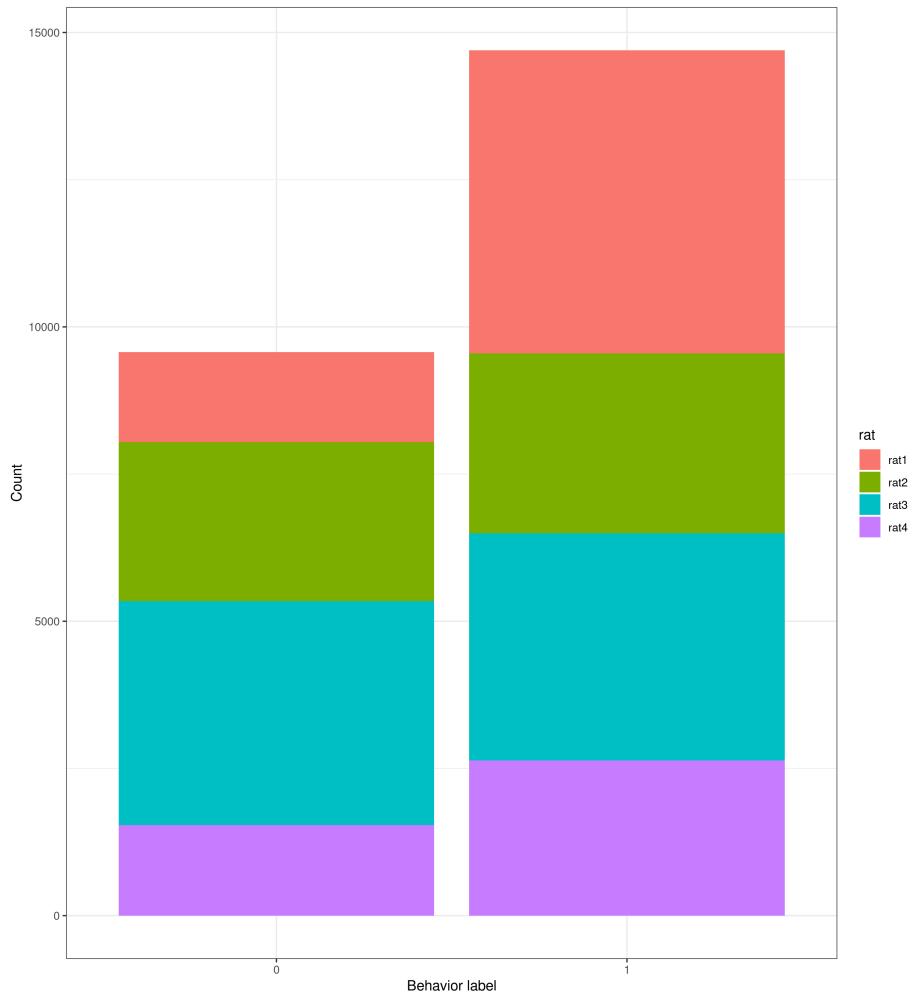


Figure 5: Proportion of labels (0 = stationary, 1 = Movement-related) over rats

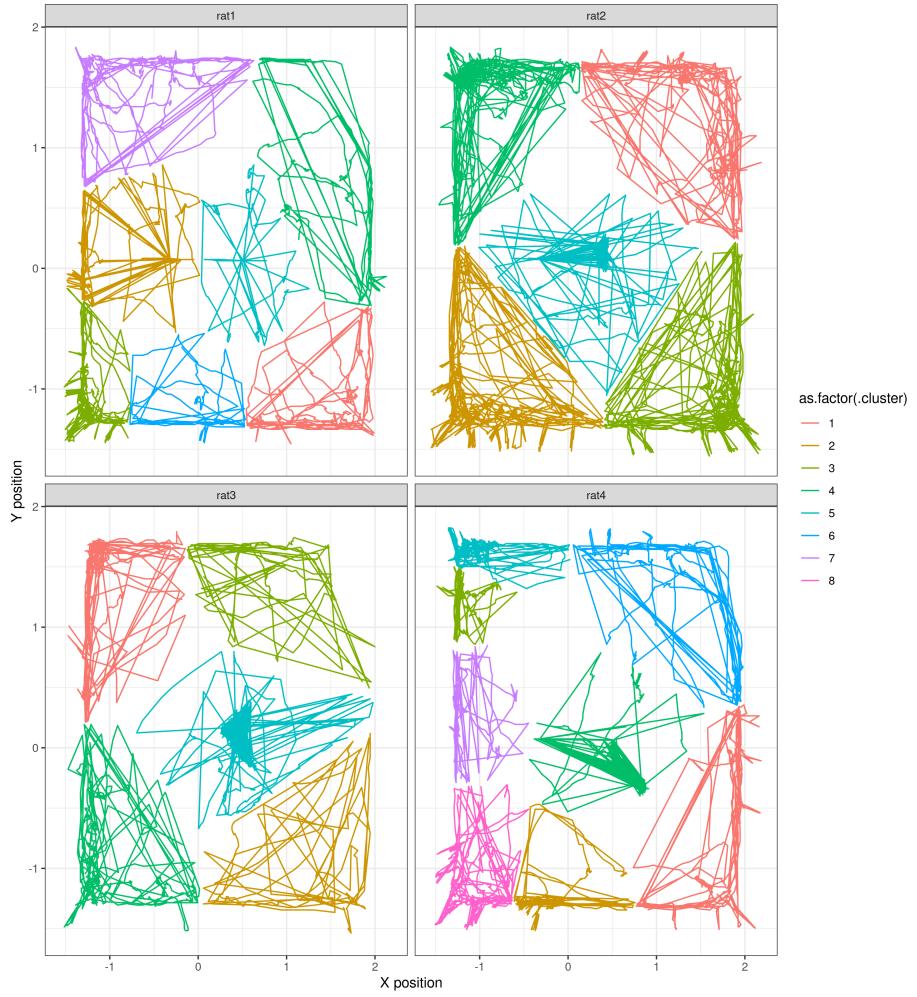


Figure 6: Graphical representation of clustering, colors indicate different clusters

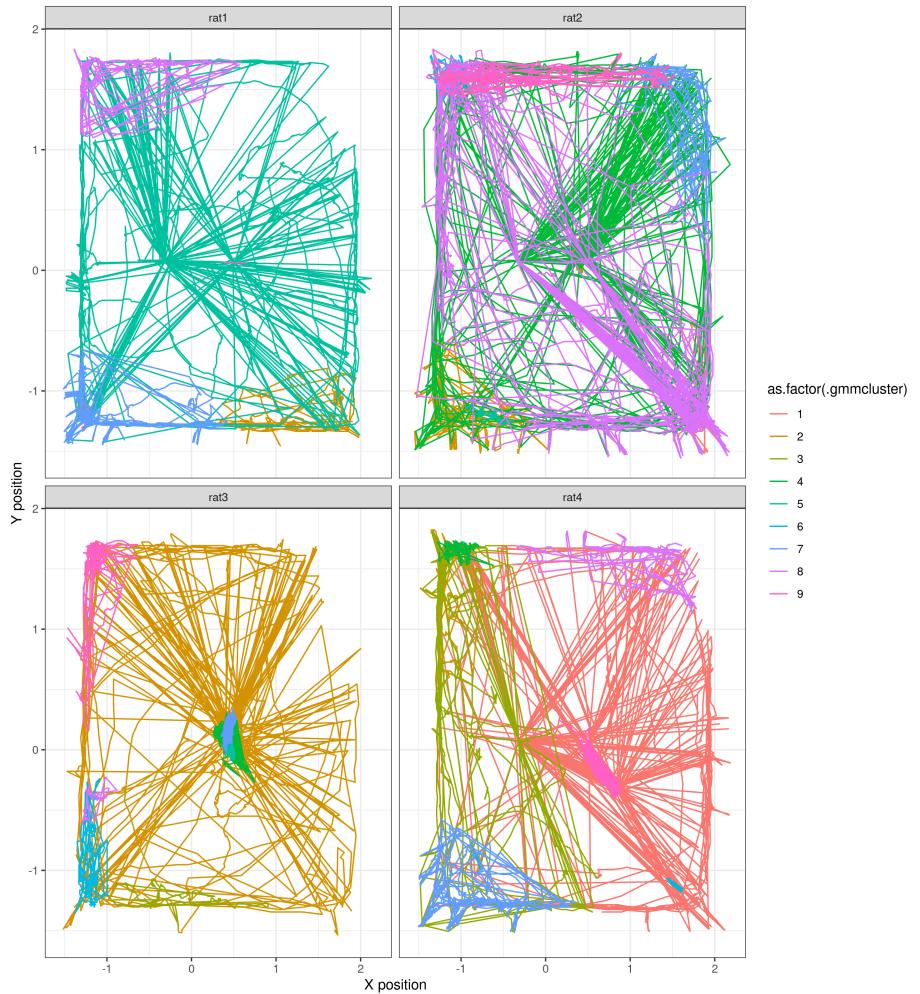


Figure 7: Visualization of GMM cluster on all rats

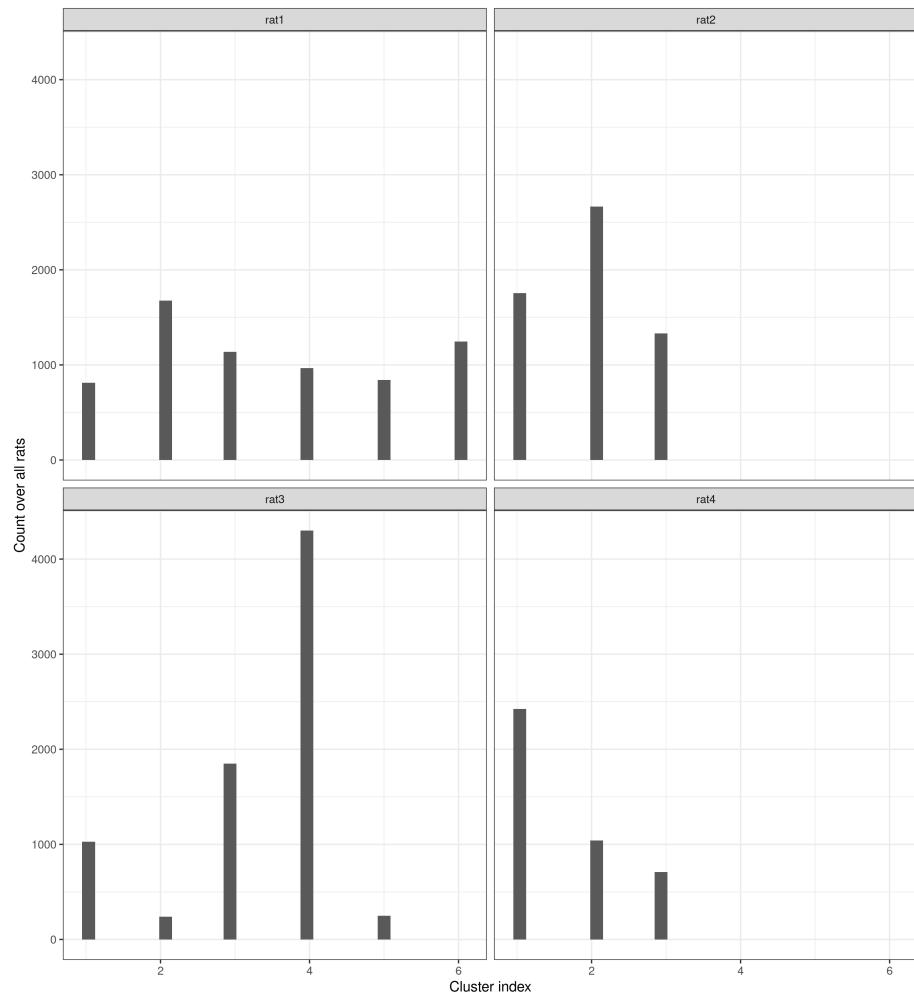


Figure 8: Visualization of GMM cluster on all rats

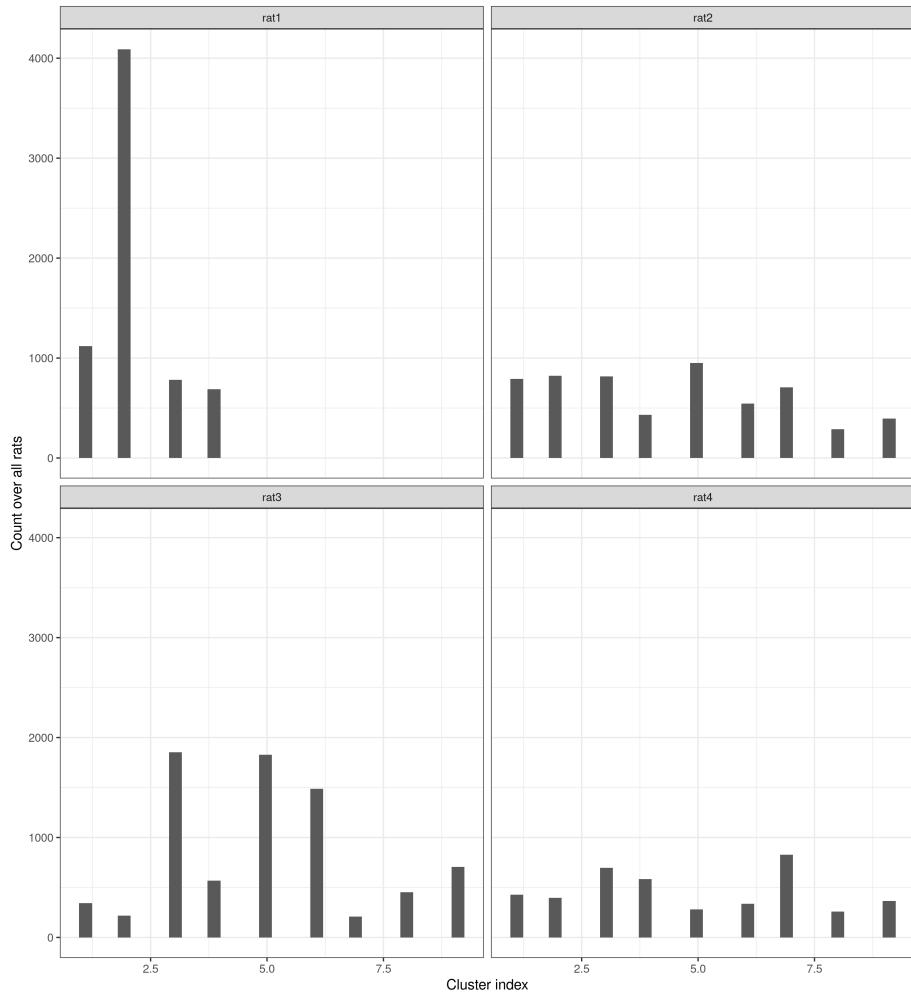


Figure 9: Visualization of GMM cluster on all rats

## References

- Chimienti, Marianna, Thomas Cornulier, Ellie Owen, Mark Bolton, Ian M. Davies, Justin M.J. Travis, and Beth E. Scott. 2016. “The Use of an Unsupervised Learning Approach for Characterizing Latent Behaviors in Accelerometer Data.” *Ecology and Evolution* 6 (3): 727–41. <https://doi.org/10.1002/ece3.1914>.
- Ellenbroek, Bart, and Jiun Youn. 2016. “Rodent Models in Neuroscience Research: Is It a Rat Race?” *Disease Models & Mechanisms* 9 (10): 1079–87. <https://doi.org/10.1242/dmm.026120>.
- Gerós, Ana, Ana Magalhães, and Paulo Aguiar. 2020. “RGB-D Rat Behavioral Dataset.” Zenodo. <https://zenodo.org/record/3636136>.
- Gris, Katsiaryna V., Jean-Philippe Coutu, and Denis Gris. 2017. “Supervised and Unsupervised Learning Technology in the Study of Rodent Behavior.” *Frontiers in Behavioral Neuroscience* 11: 141. <https://doi.org/10.3389/fnbe.2017.00141>.
- Huang, Kang, Yaning Han, Ke Chen, Hongli Pan, Wenling Yi, Xiaoxi Li, Siyuan Liu, Pengfei Wei, and Liping Wang. 2020. “Mapping Mouse Behavior with an Unsupervised Spatio-Temporal Sequence Decomposition Framework.” *Animal Behavior and Cognition*. <http://biorxiv.org/lookup/doi/10.1101/2020.09.14.295808>.
- Krakauer, John W., Asif A. Ghazanfar, Alex Gomez-Marin, Malcolm A. MacIver, and David Poeppel. 2017. “Neuroscience Needs Behavior: Correcting a Reductionist Bias.” *Neuron* 93 (3): 480–90. <https://doi.org/10.1016/j.neuron.2016.12.041>.
- Ossama, Omnia, Hoda M.O. Mokhtar, and Mohamed E. El-Sharkawi. 2011. “An Extended k-Means Technique for Clustering Moving Objects.” *Egyptian Informatics Journal* 12 (1): 45–51. <https://doi.org/10.1016/j.eij.2011.02.007>.
- Pereira, Talmo D., Joshua W. Shaevitz, and Mala Murthy. 2020. “Quantifying Behavior to Understand the Brain.” *Nature Neuroscience* 23 (12): 1537–49. <https://doi.org/10.1038/s41593-020-00734-z>.
- Sestakova, Natalia, Angelika Puzserova, Michal Kluknavsky, and Iveta Bernatova. 2013. “Determination of motor activity and anxiety-related behaviour in rodents: methodological aspects and role of nitric oxide.” *Interdisciplinary Toxicology* 6 (3): 126–35. <https://doi.org/10.2478/intox-2013-0020>.
- Shan, Haozhe, and Peggy Mason. 2020. “Unsupervised Identification of Rat Behavioral Motifs Across Timescales.” *arXiv:1707.03360 [q-Bio, Stat]*, July. <http://arxiv.org/abs/1707.03360>.
- Unger, Jakob, Mike Mansour, Marcin Kopaczka, Nina Gronloh, Marc Spehr, and Dorit Merhof. 2017. “An Unsupervised Learning Approach for Tracking

Mice in an Enclosed Area.” *BMC Bioinformatics* 18 (1): 272. <https://doi.org/10.1186/s12859-017-1681-1>.

Wiltschko, Alexander B., Matthew J. Johnson, Giuliano Iurilli, Ralph E. Peterson, Jesse M. Katon, Stan L. Pashkovski, Victoria E. Abraira, Ryan P. Adams, and Sandeep Robert Datta. 2015. “Mapping Sub-Second Structure in Mouse Behavior.” *Neuron* 88 (6): 1121–35. <https://doi.org/10.1016/j.neuron.2015.11.031>.