Mining Vehicle Trajectories to Discover Individual Significant Places: Case Study using Floating Car Data in the Paris Region

Danyang Sun1, Fabien Leurent1, and Xiaoyan Xie1

Abstract
In this study we discovered significant places in individual mobility by exploring vehicle trajectories from floating car data. The objective was to detect the geo-locations of significant places and further identify their functional types. Vehicle trajectories were first segmented into meaningful trips to recover corresponding stay points. A customized density-based clustering approach was implemented to cluster stay points into places and determine the significant ones for each individual vehicle. Next, a two-level hierarchy method was developed to identify the place types, which firstly identified the activity types by mixture model clustering on stay characteristics, and secondly discovered the place types by assessing their profiles of activity composition and frequentation. An applicational case study was conducted in the Paris region. As a result, five types of significant places were identified, including home place, work place, and three other types of secondary places. The results of the proposed method were compared with those from a commonly used rule-based identification, and showed a highly consistent matching on place recognition for the same vehicles. Overall, this study provides a large-scale instance of the study of human mobility anchors by mining passive trajectory data without prior knowledge. Such mined information can further help to understand human mobility regularities and facilitate city planning.

Mobility has been growing rapidly in recent decades. Understanding human mobility patterns has been widely acknowledged as a critical task in studying urban dynamics and facilitating sustainable development (1). With the rising implementation of location acquisition technologies such as Global Positioning System (GPS), Global System for Mobile Communications (GSM), Wi-Fi, and so on, massive datasets of location traces are becoming available and are providing fundamental knowledge with which to explore insights into people’s movements. However, such raw data merely trace geo-locations with timestamps, and therefore need to be mined to recover meaningful information from their content. One key challenge is then to recognize meaningful locations that are important to mobility makers (2). Such locations, which can be termed as “significant places,” include home place, work place, and other regular visit places which shape the mobility pattern of an individual (3). Evidence from many studies also indicates that human mobility is predictable and centered around a few base places (4, 5), while home and work places are commonly considered as the two critical base places in activity chain modeling (6) and transport planning (7). Therefore, identifying the significant places of individuals is a fundamental issue in mining mobility traces and can help us to better understand human mobility regularities.

In the previous literature, significant places have been described by different authors using different terminologies, including anchor points, core stops, meaningful locations, and so forth (3). However, a common core definition refers to the locations where people regularly stay and carry out activities. These relevant works can generally be summarized into three kinds. The first kind of work looks at activity location classification, which mainly aims to tag trajectory segments and locations with semantic meanings that are learned or trained from GPS data with labeled indicators. Corresponding methods include supervised classifiers, such as random forest models (8), and sequence inference models such as

1LVMT, UMR-T 9403, Ecole des Ponts ParisTech, UGE (ex IFSTTAR, UPEM, UPE), Champs-sur-Marne, France

Corresponding Author:
Danyang Sun, danyang.sun@enpc.fr
conditional random field (9), dynamic Bayesian network (10) and integer linear programming (2). A common issue of these studies is the requirement to have prior knowledge of the trajectory meanings. Ground truth or labels are commonly required in the classification and parameter estimation, but these are hard to obtain, especially in large-scale applications. The second group considers place inference via geographical databases. Points of interest have been widely incorporated to infer activity place types (11, 12). Other data sources containing similar information may also be explored, such as the Yellow Pages (13). However, this kind of place inference has rarely involved home place detection because of the lack of residence information in these datasets. The third kind involves mining place representations from human mobility data, in line with the interest of this study and further to the current trend of privacy protection which limits the information collected beyond anonymized traces. Ashbrook and Starner inferred significant places by clustering GPS terminated points into location clusters based on a variant of the K-means algorithm (14). A further study by Zhou et al. (15) proposed a density- and join-based clustering approach called DJ-cluster to detect significant places from GPS traces, which showed improved precision compared with the K-means. However, these two studies mainly focused on the identification of the geo-location of significant places, whereas the issue of place type inference has not been well explored. Besides these studies, mobile phone calls have been analyzed by Ahas et al. (3) and Vanhoof et al. (16) to detect home, work, and secondary places. Rule-based approaches were proposed for such an extraction based on the call-related local features of time of day and repetition over days. Chakirov and Erath developed a similar approach based on decision rules to identify activities and their primary location based on smart card payment data from the public transportation system (7). One common issue of these rule-based methods is the difficulty of determining the cut-off thresholds, which may vary from one application to another. Besides, most of the above studies modeled on a relatively small sample size with only a few selected users. The need for a method suited to large-scale data would therefore present itself when city-wide applications are considered.

The objective of this study was to identify significant places on the basis of unsupervised mining of mobility data. Vehicle trajectories traced by floating car data (FCD) were used for the exploration in the Paris region. More specifically, the geo-locations of significant places of each vehicle were first recovered from raw trajectories. After that, a two-level hierarchy approach was developed to identify the types of significant places, by analyzing the activity type based on the characteristics of each visit, and then inferring the place type per individual from its activity composition and frequentation. In contrast to previous efforts, the contribution of this work was to identify significant places on a large scale without prior knowledge of either pre-labeled data for referential training or expertise on setting specific rules for place recognition. Moreover, we aimed to identify the geo-locations and functional types of these significant places for not only the primary ones (home and work places) but also the other secondary ones.

Data Description

Mobility trajectory data are available as many kinds of sources, such as data created by cell phones, vehicle navigation devices, GPS trackers, and so forth, which are essentially similar in their data structure and representation. Vehicle trajectories traced by FCD were utilized for the exploration in this study and were aimed at showcasing the applicability of the method for GPS-based trajectory data in reality. Although the findings of this study were especially oriented to vehicle mobility, the established approaches are transferrable to other kinds of GPS trajectories in a straightforward way and can facilitate further analysis when more data are available. Our FCD were obtained by onboard GPS receivers from Coyote, a major roadway information service provider in France. These data receivers collect instantaneous states of moving vehicles along with the timestamps periodically, and include vehicle identifiers (anonymized), geo-locations (longitudes and latitudes), moving speeds, and driving headings (azimuth). The raw dataset was organized as sequences of logs, each representing a vehicle trace of the above information. The recording frequency was around 30s to 60s, depending on the device configuration of different models.

FCD in the Paris region were analyzed in this study with a time span of 14 days in February 2019. Trajectories of 168,308 unique vehicles with a total of over 15 million logs were obtained within the Paris region during the time period. However, with computation efficiency in mind, a pre-selection was performed to downsize the dataset. Consequently, 10,000 vehicles with a frequent usage, that is, defined as having been used on at least seven distinct days, were finally adopted in the following analysis. It should be noted that this does not affect the methodological findings in a meaningful way as the significant places were identified at the level of each individual vehicle.

Significant-Place Geo-Location Detection

Trajectories collected by FCD are represented as sequences of timestamped traces with no direct semantic meanings. This section describes the mining task conducted to process the trajectory and recover the geo-locations of significant places. Here, we define some
preliminary notions: (1) a trip indicates a segment of a trajectory which a vehicle makes for a certain purpose of movement; (2) the temporal interval between two trips indicates the time duration of an activity; (3) a stay point indicates the geo-position at the end of a trip; and (4) a place indicates a space where people carry out activities, the geo-location of which can be represented by a cluster of sufficiently adjacent stay points.

Stay Point Detection

The first step of mining was to sequence the trajectories of each vehicle and segment them into meaningful trips. Many studies have been carried out on such trajectory sequencing on GPS traces, but most of them focus on personal trips based on mobile phone data \((17, 18)\). Only a limited amount of work has been done on a vehicle basis \((19, 20)\). Generally, the time interval between two succeeding points is most widely adopted as the decisive feature for identifying trip ends, with other criteria incorporated by some studies including distance and average interval speed. However, thresholds of those criteria varied from one to another according to the data source and local context. In this study, we developed a trip identification algorithm, extended from our previous study \((21)\), by adopting two main criteria—the time interval \(t\) and the interval distance ratio \(i\)—which leverage both distance and speed information recorded in the data with the formula

\[
i = \frac{\text{observed distance}}{\text{expected distance}} = \frac{\text{distance} (P_a \text{ to } P_b)}{0.5*(V_a + V_b)*t}
\]

where distance \((P_a \text{ to } P_b)\) indicates the planar distance between point \(a\) and point \(b\) of an interval, and \(V_a\) and \(V_b\) indicate the recorded instantaneous speed of the two points, respectively. The process of the algorithm is illustrated in Figure 1. Stay points are the stopping places where a trip terminates for an activity. It can be inferred as having been made during an interval with time duration exceeding 90 s and an insignificant geographical movement. However, many confounding scenarios may exist as a result of unstable signal transmission and other asynchronous triggers such as engine-off events. The overall hypothesis was that for each interval, the vehicles with a trip end should have a much lower observed distance than that expected from those without a trip end. A threshold of \(i = 0.2\) was derived by inspecting the distribution and conducting a kernel density estimation on all the data to segment the mixture. For qualified intervals, the geo-position of \(P_a\) was used to approximate the geo-location of the true stay point as it was the last recorded point of the trip. It should be noted that although not all trips could be detected, this algorithm aimed to capture those with major trip-end activities and without involving too many “false positives.” As a result, a new dataset of detected stay points was generated along with corresponding descriptive features including geo-locations, arrival times, and activity durations. A total of 307,623 trips were identified for the 10,000 vehicles.

Stay Point Clustering to Extract Significant Places

Stay points at the end of trips can indicate the place where activities are carried out. However, because of the complicated nature of geo-locations, stay points with slightly different geo-positions may correspond to the same place. The second step of mining was therefore conducted to cluster those adjacent points into spatial clusters, using their centroids to represent the geo-locations of meaningful places \((22)\). The process is illustrated in Figure 2. In the meantime, significant places can also be extracted by selecting spatial clusters with multiple visits by the same vehicle. To achieve the above goals, density-based clustering methods can be employed to detect spatial clusters based on geo-coordinates \((16, 22)\). The widely used Density-based Spatial Clustering of Applications with Noise (DBSCAN) was adopted in this study. Such a method requires two input parameters,
namely distance threshold (eps) and minimum number of points (minPts) to form a cluster. Considering the semantic interpretations, the eps was set as 150 m and the minPts was set as 2. One common issue of this algorithm is that some large clusters may be formed by straight cluster chains if many places are densely connected in urban settings. As well as adjusting the thresholds, we customized the clustering process by running it for individual vehicles separately to avoid such chaining cases; the results were visually inspected on their geographical distributions and fitted well in our application. As a result, the geo-locations of 44,882 significant places were identified with a vehicle average of 4.48.

It should also be noted that places were treated separately for different vehicles by bundling place identifiers with vehicle identifiers. This is because a single place can have different meanings for different visitors; for example, a shopping mall may represent the work place for vehicle user 1, while just playing the role of an entertainment stop for others.

**Significant-Place Type Identification**

Places may play different roles in people’s life, with some places visited more frequently and others less. Investigating visit characteristics at places can help to differentiate their types so as to understand individual mobility regularities. In this section, the significant places extracted earlier were examined for each vehicle with the aim of identifying their functional types such as home place, work place, and other types of frequent visits.

**Method: A Two-Level Identification**

In previous works, features related to significant-place recognition can generally be summarized into two groups: (1) temporal circumstances, including time of the day, day of the week (16, 23) and activity duration when visiting the place; (2) repetition pattern: frequency on distinct days of visits to the place (16). Additionally, it is reasonable to assume that activities occurring at a place would strongly indicate the type of place (9). In this study, a two-level identification method was proposed, which analyzed the temporal circumstances at Level 1 to identify the activity type of each single stay and then identified the place type at Level 2, based on its activity composition and repetition on different days. The concept of the hierarchical framework is illustrated in Figure 3.

**Significant-Place Type Identification Based on Activity Composition and Frequentation.** The functional types of places are characterized by their related activities. For a place i of individual k, an activity profile can be built by counting the frequency of different activity types j over distinct days, which can be formed as a vector of counts $F_{ik} = (f_{ik0}, \ldots, f_{ikN})$. To exclude the effects of vehicles not necessarily being used every day or being used equally frequently, each count $f_{ikj}$ was further adjusted to a conditional relative frequency by dividing it by the...
Such a conditional relative frequency $f_{ikj}$ reflects the likelihood of the activity being carried out in the place as a daily average. This adjusting process also helps to weigh each feature dimension equally in the following clustering process.

The $K$-means clustering algorithm was employed to partition these profiles to detect different place types. The optimal number $k$ of clusters is commonly recognized as a subjective issue and depends on the nature of the data used for partitioning. In this study, the two widely used methods, the elbow method and the average silhouette method, were employed to determine the optimum cluster number. By comparing the suggested good candidates from the two analyses, a final choice of the optimal number of $k$ can be determined.

**Results of Activity Type Identification**

GMM clustering was run using the Scikit-learn implementation. The number of mixture components was determined as four on weekdays and three on weekends,
as visually suggested from the density distribution plots (shown in Figures 4a and 5a). The stays were therefore partitioned into four clusters for weekday scenarios and three for weekend scenarios, as shown in different colors in the scatter plots in Figures 4b and 5b. Considering the cyclical nature of time, activities that were after midnight but before dawn were considered as belonging to the previous day. By comparing the attribute characteristics, the activities can be characterized into four and three types respectively, along with descriptive statistics summarized in Table 1. Both the weekday and weekend scenarios were found with an activity type with a long duration and visiting time in the late part of a day, implying those trips were journeys home at the end of a day. Compared with weekends, there was one more type detected on weekdays, that with a long duration and visiting time in the early part of day. Such a type can be interpreted as being related to work, which is consistent with the finding that these trips were only detected in major numbers on weekdays. The other two types, early-day short and late-short activities, were found for both weekdays and weekends, but with the duration slightly longer and the visiting time a bit later on weekdays.

### Results of Significant-Place Type Identification

Seven place clusters (c0–c6) were drawn from the $K$-means clustering on their activity profiles. The occurrence likelihoods of activities ($f_{ijk}, \ldots, f_{ijk}$) are displayed for each place cluster by heatmaps in Figure 6. The color gradient in the heatmap indicates the mean value of the likelihoods in each cluster. There was a total

#### Table 1. Identified Activity Types and Corresponding Statistics

| Identified activity types | Visiting time of day (s) | Activity duration (s) |
|---------------------------|--------------------------|------------------------|
|                           | Mean         | SD          | Mean      | SD         |
| Weekday (wkd)             |              |             |           |            |
| Early-day short           | 36,436       | 7,948       | 4,249     | 5,187      |
| Early-day long            | 31,526       | 6,036       | 33,738    | 6,608      |
| Late-day short            | 60,287       | 8,853       | 4,842     | 5,864      |
| Late-day long             | 65,981       | 15,084      | 49,481    | 13,438     |
| Weekend (wkn)             |              |             |           |            |
| Early-day short           | 37,467       | 9,268       | 5,218     | 8,267      |
| Late-day short            | 60,613       | 9,476       | 6,215     | 7,490      |
| Late-day long             | 56,883       | 23,357      | 54,878    | 15,029     |

Note: SD = standard deviation. Seconds for time of day start as 0 from midnight.
of seven activity types \((j = 0, 1, \ldots, 6)\) according to the activity type identification. The qualitative definition of “frequent” is a subjective issue and we assumed that a likelihood of over 0.33 could be regarded as a frequent pattern as it implies a visit at least every 3 days. By comparing the corresponding prominent characteristics, the places were characterized into five types, as described below:

- **Home places \((c0 \& c5)\)**: These two clusters shared similar patterns by showing a significantly high chance of late-day long activities on both weekdays and weekends, which is a typical pattern for home places or residence places. Compared with \(c0\), places in cluster \(c5\) also included more of the other types of short activities with a greater emphasis during the second half-day. This might indicate that vehicles pertaining to those home places were more frequently used between the home and other places rather than simply commuting behaviors.

- **Work places \((c2 \& c6)\)**: Places in cluster \(c6\) showed a comparatively high likelihood of early-day long activities during weekdays, which implies that their major place occupations are related to long working stays throughout the day. Cluster \(c2\) showed frequent visits for short activities in the morning and afternoon only on weekdays. This corresponds to the pattern of work places for those who leave their work places during the noon break for another purpose such as lunch at home or at restaurants. It should be noted that these work-related activity likelihoods were not found to be as high as close to 1. However, such a finding may be consistent with people not necessarily driving cars to work every day.

- **Secondary places I—Weekend frequently visited places with late-day short stay \((c4)\)**: This cluster showed a pattern of frequent short visits in the second half-day on weekends, which may represent secondary places for typical weekend activities, such as frequently visited shops, recreation spaces, or favorite meet-up places.

- **Secondary places II—Weekend moderately frequently visited places with early-day short stay \((c3)\)**: Places in this cluster encountered a few more visits during the first half-day periods on weekends, which can be interpreted as secondary places such as markets or bakeries.

- **Secondary places III—Weekday less frequently visited places \((c1)\)**: Places in this cluster did not show any significantly frequent visits, which may indicate places that individuals come to and visit from time to time (at least twice) but with no significant regularity. Overall, the activity patterns were more distributed among weekdays.

Consequently, 8,111 out of the 10,000 vehicles studied were identified with home places and 4,295 vehicles were identified with work places. Such a portion of recognition is not high, especially with work place detection, but is consistent with the home place normally displaying more regular patterns on a vehicle usage basis while patterns of work places do not, among individuals. The analysis timeframe of 14 days was also a limitation restricting the detection of frequency patterns.

**Comparison with Rule-Based Identification**

Validation of such explorations is a common problem in current research into mobility pattern mining because of the lack of ground truth for the evaluation. Within this restriction, the results using the proposed significant-place identification method were compared with those from a benchmark method: rule-based extraction, which was widely used in previous studies \((3, 7, 16)\). The comparison was mainly on home and work places because of their ability to be characterized with explicit criteria. The criteria for the decision rules were set for the features directly at the place level, described as below:

- **Home place**: (1) High probability \((p > 0.5)\) of being the first/end place visited during a day sequence. (2) High attendance among distinct days \((>80\%)\) when the vehicle is used.

- **Work place**: (1) More daytime \((8–18h)\) activities at the places; (2) Highest attendance during

---

**Figure 6.** Identified place clusters and corresponding activity frequentation patterns.
weekdays (>60%); (3) Long average activity duration (>2 h).

Through the rule-based extraction, 5,822 vehicles were found with home places and only 1,375 vehicles among them were detected with work places. The use of cut-off rules is limited in detecting diverse patterns, but we can assume that the extracted parts are representative of home and work places. The comparison between our method and the rule-based method was based on the 5,822 vehicles by considering the type labels of the rule-based method as the base reference. The comparison results are summarized in Table 2, with each cell representing the number of places labeled in the two methods. All numbers were counted on the unit of significant places identified for each individual. As a result, our method showed a high consistency ratio against the rule-based method, with 96% for home places, 90% for work places and 89% for the secondary places.

### Conclusion and Discussion

This study proposed a methodological approach to identify individual significant places by mining vehicle trajectories from FCD. A case study was conducted in the Paris region for an experiment. Meaningful trips were segmented from raw trajectories to recover stay points. Customized density-based clustering by DBSCAN was implemented to cluster stay points into places and extract the geo-locations of significant places for each individual vehicle. Next, a two-level hierarchy method was developed to identify the types of these significant places. At the first level of the process, it identified the types of each stay activity by GMM clustering based on the visit time of day and the activity duration. At the second level, it found the types of places by clustering their profiles by occurrence likelihoods with different activities. As a result, five types of significant places were derived, including home place, work place, and three other types of secondary places. Most of the vehicles analyzed were detected with home places while around half of them were found with work places based on frequent vehicle usage. The results of the proposed method were also compared with those from the commonly used rule-based extraction method, and showed a highly consistent matching. Moreover, based on unsupervised mining, our proposed method was less dependent on the expertise of rules and more applicable for detecting places displaying more diverse situations. Although the findings of this paper were based on the vehicle mobility because of the nature of the data, the proposed methodology is transferable to other kinds of GPS-based trajectory data (by cell phone, GPS tracker, etc.) in a straightforward way.

Overall, this research provides a large-scale instance of identification of significant places by both geo-locations and types without prior knowledge, which shows its applicability to human geography in a cost-effective way by leveraging digital trajectories. The information mined can be used as a basis for further studies of human mobility regularity and predictability, and thus be of benefit to future city planning. The results were restricted because of a limited 14-day period of available data. Future work could be done on a broader timespan to explore the place visiting frequency on a larger scale, such as monthly regularities. Data from other modes of mobility may be explored together to obtain a more complete view of individual mobilities. Complementary sources of data, such as land use and geographical reference data may also be incorporated to facilitate the place identification or for comparison with the results for validation purposes.

### Acknowledgments

The authors would like to thank the Geolytics project, with partners IT4PME and Coyote, for the data and their assistance, and the ENPC-IdFM Chair for facilitating this research.

### Author Contributions

The authors confirm their contributions to the paper as follows: study conception and design: D. Sun, F. Leurent, X. Xie; data collection: D. Sun, F. Leurent, X. Xie; analysis and interpretation of results: D. Sun; draft manuscript preparation: D. Sun, F. Leurent, X. Xie. All authors reviewed the results and approved the final version of the manuscript.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported by VINCI ParisTech Environment Research Lab at Ecole des Ponts ParisTech.

### References

1. Zheng, Y., L. Capra, O. Wolfson, and H. Yang. Urban Computing: Concepts, Methodologies, and Applications.
13. Cao, X., G. Cong, and C. S. Jensen. Mining Significant Semantic Locations from GPS Data. Proceedings of the VLDB Endowment, Vol. 3, No. 1–2, 2010, pp. 1009–1020.

14. Ashbrook, D., and T. Starner. Using GPS to Learn Significant Locations and Predict Movement across Multiple Users. Personal and Ubiquitous Computing, Vol. 7, No. 5, 2003, pp. 275–286.

15. Zhou, C., D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen. Discovering Personally Meaningful Places: An Interactive Clustering Approach. ACM Transactions on Information Systems (TOIS), Vol. 25, No. 3, 2007, p. 12-es.

16. Vanhoof, M., F. Reis, T. Ploetz, and Z. Smoreda. Assessing the Quality of Home Detection from Mobile Phone Data for Official Statistics. Journal of Official Statistics, Vol. 34, No. 4, 2018, pp. 935–960.

17. Gong, L., H. Sato, T. Yamamoto, T. Miwa, and T. Morikawa. Identification of Activity Stop Locations in GPS Trajectories by Density-Based Clustering Method Combined with Support Vector Machines. Journal of Modern Transportation, Vol. 23, No. 3, 2015, pp. 202–213.

18. Gong, L., T. Morikawa, T. Yamamoto, and H. Sato. Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies. Procedia-Social and Behavioral Sciences, Vol. 138, Supplement C, 2014, pp. 557–565.

19. Chen, H., C. Yang, and X. Xu. Clustering Vehicle Temporal and Spatial Travel Behavior using License Plate Recognition Data. Journal of Advanced Transportation, Vol. 2017, No. 7, 2017, pp. 1–14.

20. Lin, K., Z. Xu, M. Qiu, X. Wang, and T. Han. Noise Filtering, Trajectory Compression and Trajectory Segmentation on GPS Data. Proc., 2016 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, IEEE, New York, 2016, pp. 490–495.

21. Sun, D., F. Leurent, and X. Xie. Floating Car Data Mining: Identifying Vehicle Types on the Basis of Daily Usage Patterns. Transportation Research Procedia, Vol. 47, 2020, pp. 147–154.

22. Ye, Y., Y. Zheng, Y. Chen, J. Feng, and X. Xie. Mining Individual Life Pattern Based on Location History. Proc., 2009 10th International Conference on Mobile Data Management: Systems, Services and Middleware, Taipei, Taiwan, IEEE, New York, 2009, pp. 1–10.

23. Tongsinoot, L., and V. Muangsin. Exploring Home and Work Locations in a City from Mobile Phone Data. Proc., 2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Bangkok, Thailand, IEEE, New York, 2017, pp. 123–129.

24. Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Numerical Recipes: The Art of Scientific Computing, 3rd ed. Cambridge University Press, New York, NY, 2007.

25. Scikit-Learn 0.23.1 Documentation. Gaussian Mixture Models. https://scikit-learn.org/stable/modules/mixture.html #gmm. Accessed July 31, 2020.