From supervised to unsupervised learning - Structuring the core components of understanding

Katharina Bata
Karlsruhe Institute of Technology, Germany
Katharina.bata@kit.edu

With the increasing public relevance of machine learning, the need for corresponding educational opportunities is also growing. While theoretical concepts for structuring learning content already exist for supervised learning, there is no comparable basis for unsupervised learning. This paper examines the extent to which the so-called model concept, a theoretical model to structure the core components of understanding of supervised learning, can be transferred to unsupervised learning. Using the example of k-means cluster analysis, it is shown that the basic structure of the model concept, the so-called facets, is largely transferable, even if individual core components of understanding need to be re-differentiated in terms of content. The results provide a theoretical basis for the development of learning objectives and teaching materials for unsupervised learning and open up further questions regarding the implementation and empirical validation of the proposed structure.

THEORETICAL BACKGROUND AND RESEARCH QUESTION

With the increasing relevance of machine learning (ML) in everyday life, for example, through chatbots or recommender systems the topic is gaining attention at all levels of the education system (UNESCO, 2024). ML lies at the intersection of computer science and mathematics including statistics and probability theory, which means that the underlying concepts of the techniques are sometimes very complex. In addition, the fruitful application of the techniques also requires knowledge of the problem and the underlying data, as well as potential social and ethical issues. A broader societal goal should be that students of all ages acquire foundational knowledge that enables them to recognize and critically assess systems created through ML (Tedre, 2021). For university teaching in particular, there are further goals, depending on the degree subject, which focus more on the mathematical and technical subtleties of the processes (King, 2019; Kandlhofer, 2016; Shaprio, 2018).

Research on ML teaching and learning for different target groups gained traction in recent years. Reviews in the field of K12 education show that ML concepts, algorithms, and tasks can be captured by students of different age groups (Martins & Gresse von Wangenheim, 2022; Marques et al., 2020). Furthermore, reviews show that many of the publications deal specifically with the development of pedagogical approaches, curricula, and tools to support ML education in K-12 settings (Sanusi et al., 2023). With regard to the learning processes. Also at the university level, discussions about how ML should be integrated into curricula are ongoing (Shapiro, 2018) and there are a lot of creative best practices to teach ML for different target groups (Garcia-Algarra, 2020; Huppenkothen & Eadie 2020).

Two popular approaches in ML are supervised learning and unsupervised learning. Compared to supervised learning, unsupervised learning is less commonly addressed in educational contexts (Marques et al., 2020), and consequently, little research exists on how to teach it - for example, through the visualization of models (Fuchs, 2019). From the perspective of statistics and data science education, however, unsupervised learning is gaining increasing attention as a learning subject, since the mathematical methods underlying ML techniques are based on statistical procedures and are used to process data and extract information (Biehler et al., 2022). Fundamental concepts from statistics education, such as data, center, and covariation (Garfield et al., 2008), are also relevant in unsupervised ML techniques, for instance when identifying associations in association analysis or when calculating clusters in cluster analysis.

In the context of supervised learning, the so-called “model concept” has been developed (Bata, 2025). This concept represents and structures the core components of understanding for models created with ML techniques (ML models). Since the model concept has proven to be empirically feasible and thus provides a foundation for the development of learning objectives and teaching materials for learning objects from supervised learning, this paper investigates whether the model concept can be transferred to the subject of unsupervised learning.

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*. © ISI/IASE.

SUPERVISED VS. UNSUPERVISED LEARNING

Supervised learning is based on a dataset that consists of independent features and a corresponding (potentially dependent) feature, sometimes called the label. The goal of supervised learning is to derive a mapping from the data that assigns a label to any combination of the independent features. A classic example in educational research would be the qualitative coding of a dataset using a given category system, based on a large set of already coded data. The mapping determined by supervised learning could then be applied to new data to identify relevant parts of a statement and assign them to the "correct" category.

In contrast, the dataset used in unsupervised learning contains no labels. The goal is to identify unknown patterns or groups within such a dataset. Compared to the example above, a task suitable for unsupervised learning would involve qualitatively coding a dataset without a given category system and based on data that have not been pre-coded.

Both supervised and unsupervised learning rely on mathematical analysis of the underlying data. Examples for supervised learning are decision trees or regression functions (Bishop, 2006). One example for unsupervised learning is the k-means cluster analysis (Bishop, 2006). The goal of k-means clustering is to determine whether a set of data points can be meaningfully divided into k so called clusters. Clusters are groups with shared characteristics. Starting from randomly chosen cluster centroids, clusters are assigned by allocating each data point to the nearest centroid. New centroids are then computed from the resulting clusters. By iteratively repeating this process, the centroids typically stabilize, ultimately determining the final clusters. The result of the analysis is a model of the structure abstracted from the data, where each data point is associated with cluster information as a newly defined feature.

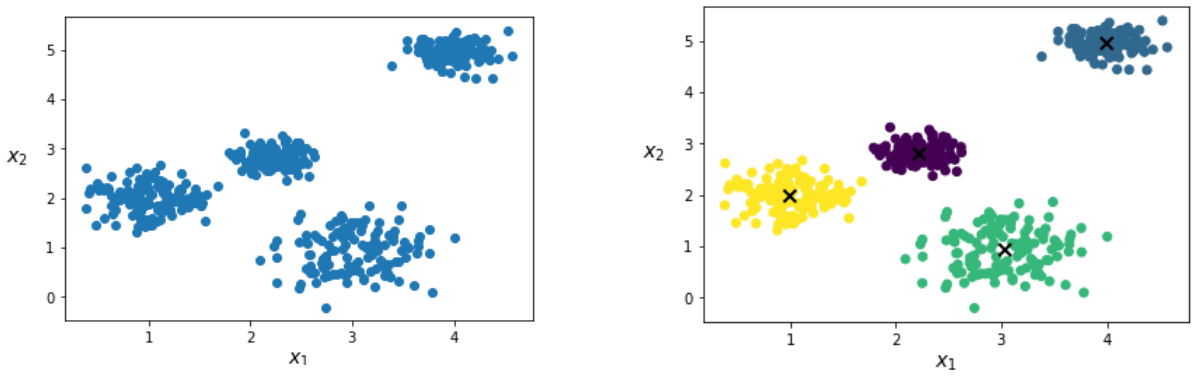

Figure 1. Example for a k-means clustering.

CORE COMPONENTS OF UNDERSTANDING IN UNSUPERVISED LEARNING

The model concept (Bata, 2025) is a theoretical construct used to represent and structure conceptual knowledge about ML models. At its core are four so-called facets: data, assignment, quality, and usage. These facets, based on Mahr's general model theory (Mahr, 2008), denote the conceptual focus with which the ML model is analyzed in each case. Mahr's theory states that the model becomes a model through three different perspectives, which all need to be meaningful. One perspective is that of the model object itself, one is the production perspective, which deals with the connection between the source object and the model object, and the last is the application perspective, which focuses on the applications of the model object to gain information.

For each facet, specific core components of understanding are formulated. According to Drollinger-Vetter (2011), core components of understanding are the components of a concept that must be grasped in order to understand the concept as a whole. As in the model theory of Mahr (2008), the facets are interdependent, which makes the inclusion of transfer areas necessary. These transfer areas contain core components of understanding that span multiple facets or form connections between them. Figure 2 shows a visual representation of the model concept, in which the different facets are positioned

at the corners, with the usage facet shown as an overarching facet. The transfer areas connect two facets each, while the central transfer area links three facets. These surfaces provide an overview of all key domains of conceptual knowledge and can also be used to note down individual core components of understanding or to make a knowledge state visually explicit (Bata, 2025).

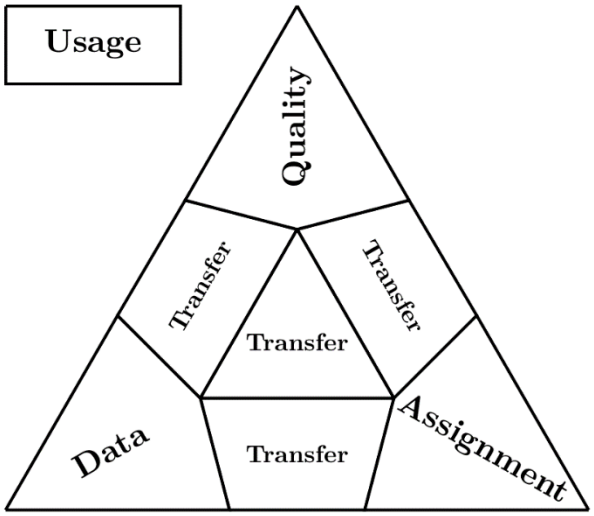

Figure 2. The model concept und its facets (Bata, 2025).

The central question of this paper is whether these four facets, data, assignment, quality, and usage, are sufficient to cover the core components of understanding for unsupervised learning as well. The transfer areas retain their original function, continuing to represent those core components of understanding that are linked to more than one facet.

In its original formulation, the data facet as the equivalent of the production perspective in Mahr (2008) addresses the data basis of the model. Since data serve a similar foundational role in both supervised and unsupervised learning, this facet is also relevant for unsupervised learning. Core components of understanding such as ‘A model depends on data’ or ‘Data can contain errors and outliers’ (Bata, 2025) can be formulated for unsupervised learning as well.

The usage facet and the quality facet, which represent the application perspective of Mahr (2008) address the use of the model for a specific question as well as the model properties at the technical level and with regard to usage. Since both the quality of the model and its usability are important criteria, these two facets are also reasonably transferable. However, the specific core components of understanding in these facets will likely differ from those in the original model concept. This is due, for instance, to differences in quality metrics: in supervised learning, quality is typically assessed based on the deviation between the model’s assigned label and the actual label (e.g., accuracy rate). In unsupervised learning, where no such label exists, different metrics must be considered.

The assignment facet in the original model concept focuses on the model’s property of assigning features to labels. Since unsupervised learning lacks explicit labels, this element might seem missing. However, identifying unknown patterns or groups in the data also involves making these patterns visible. Naming these patterns or groups, such as the clusters found through k-means, represents an assignment of overarching properties based on features and justifies the continued relevance of the assignment facet. As with the use and quality facets, at least some of the core components of understanding within the assignment facet will differ from the original model concept. General core components such as 'a model assigns' or 'the assignment follows a prescriptive process' may still apply, while core components describing different supervised learning procedures (e.g., 'partitioning the datasets by a function') will obviously not.

DISCUSSION

This paper has provided a foundation for transferring the model concept (Bata, 2025) to the domain of unsupervised learning. As exemplified in the discussion above, the model concept can support the formulation and structuring of core components of understanding. The resulting structure can be used to identify learning objectives and to develop supporting teaching-learning materials (Hußmann & Prediger, 2016).

In actual teaching design, further questions arise from the proposed transfer. Examples for those questions concerns connections to students' prior knowledge or ways to let the learners build the mentioned mental model. This issue is relevant whether unsupervised learning is taught in isolation, e.g., as a workshop in schools (Bata & Frank, 2025), or as part of a standard introductory course on ML in higher education. In the latter case, another question is whether the conceptual similarity between supervised and unsupervised learning, when taught in each after another, acts more as a bridge or a barrier to understanding.

A second question concerns the robustness of a theoretically derived set of core components of understanding. Regardless of the teaching context, this question points to the need for empirical validation of the structure and potential enrichment of the model with additional core components of understanding.

REFERENCES

Bata, K. (2025). *Maschinelles Lernen lernen: Entwicklung und Erforschung einer Lehr-Lernumgebung in den Ingenieurwissenschaften* [Learning machine learning – Development and investigation of a teaching-learning environment in engineering sciences]. Springer Spektrum. https://doi.org/10.1007/978-3-658-47458-4

Bata, K., & Frank, M. (2025). Unsupervised machine learning as learning content in lower secondary school. In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025), Salzburg, Austria*. https://openreview.net/forum?id=FxrQdRUC6y

Biehler, R., De Veaux, R., Engel, J., Kazak, S., & Frischemeier, D. (2022). *Editorial: Research on data science education*. *Statistics Education Research Journal, 21*(2). https://doi.org/10.52041/serj.v21i2.606

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit: Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht* [Core components of understanding and structural clarity: Subject-matter quality in guiding mathematical understanding in teaching]. Waxmann.

Fuchs, J., Isenberg, P., Bezerianos, A., Miller, M., & Keim, D. (2019). EduClust – A Visualization Application for Teaching Clustering Algorithms. In M. Tarini & É. Galin (Eds.), *Eurographics 2019 – Education Papers* (pp. 9–16). The Eurographics Association. https://doi.org/10.2312/eged.20191023

Garcia-Algarra, J. (2020). Introductory machine learning for non-STEM students. In P. Steinbach, H. Seibold, & O. Guhr (Eds.), *Proceedings of the First Teaching Machine Learning and Artificial Intelligence Workshop* (pp. 7–10). PMLR. https://proceedings.mlr.press/v141/garcia-algarra21a.html

Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* Springer. https://doi.org/10.1007/978-1-4020-8383-9

Huppenkothen, D., & Eadie, G. (2020). Teaching the foundations of machine learning with candy. In P. Steinbach, H. Seibold, & O. Guhr (Eds.), *Proceedings of the First Teaching Machine Learning and Artificial Intelligence Workshop* (pp. 29–35). PMLR. https://proceedings.mlr.press/v141/huppenkothen21a.html

Hußmann, S., & Prediger, S. (2016). Specifying and structuring mathematical topics. *Journal für Mathematik-Didaktik, 37*(S1), 33–67. https://doi.org/10.1007/s13138-016-0102-8

Kandlhofer, M., Hirschmugl-Gaisch, S., Huber, P., & Steinbauer, G. (2016). Artificial intelligence and computer science in education: From kindergarten to university. In *2016 Frontiers in Education Conference (FIE)* (pp. 1–5). IEEE. https://doi.org/10.1109/FIE.2016.7757493

King, S. O. (2019). How electrical engineering and computer engineering departments are preparing undergraduate students for the new big data, machine learning, and AI paradigm: A three-model overview. In *2019 IEEE Global Engineering Education Conference (EDUCON)* (pp. 352–356). IEEE. https://doi.org/10.1109/EDUCON.2019.8725072

Mahr, B. (2008). Ein Modell des Modellseins. Ein Beitrag zur Aufklärung des Modellbegriffs [A model of being a model: A contribution to clarifying the concept of model]. In U. Dirks & E. Knobloch (Eds.), *Modelle* [Models] (pp. 187–218). Peter Lang.

Marques, L. S., Gresse von Wangenheim, C., & Hauck, J. C. R. (2020). Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education, 19*(2), 283–321. https://doi.org/10.15388/infedu.2020.14

Martins, R. M., & Gresse von Wangenheim, C. (2023). Findings on teaching machine learning in high school: A ten-year systematic literature review. *Informatics in Education, 22*(3), 421–440. https://doi.org/10.15388/infedu.2023.18

Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2023). A systematic review of teaching and learning machine learning in K–12 education. *Education and Information Technologies, 28*(5), 5967–5997. https://doi.org/10.1007/s10639-022-11416-7

Shapiro, R. B., Fiebrink, R., & Norvig, P. (2018). How machine learning impacts the undergraduate computing curriculum. *Communications of the ACM, 61*(11), 27–29. https://doi.org/10.1145/3276742

Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). Teaching machine learning in K–12 classrooms: Pedagogical and technological trajectories for artificial intelligence education. *IEEE Access, 9*, 110558–110572. https://doi.org/10.1109/ACCESS.2021.3097962

UNESCO. (2024). *AI competency framework for students*. https://doi.org/10.54675/JKJB9835