

# An Overview of Unsupervised Learning

Machine learning algorithms that are built to predict a target or response value (either real-valued or categorical) fall under the heading of supervised learning and include Linear and Logistic Regression, k-Nearest Neighbors, Support Vector Machines, Decision Tree-based models, and Neural Networks. Since all these tools require a well-defined y-value to predict, they can only be used with a relatively small set of data science problems.

With unsupervised learning problems, we begin with a dataset that does not have this target value. In this case, our goals will usually be related to simplifying or summarizing the data. Since we lack a dimension or variable that has the special status of being the target, we can't really talk about how well our model is doing as clearly as is possible with supervised models. We are not able to do cross-validation, so at some level choosing what's "good enough" will always be a judgment call in unsupervised learning. That said, it can be very useful to build a "big picture" representation of a dataset using unsupervised learning.