

# Machine Learning - Unsupervised

---

So far what you have seen is making the machine learn to find out the solution to our target. In regression, we train the machine to predict a future value. In classification, we train the machine to classify an unknown object in one of the categories defined by us. In short, we have been training machines so that it can predict Y for our data X. Given a huge data set and not estimating the categories, it would be difficult for us to train the machine using supervised learning. What if the machine can look up and analyze the big data running into several Gigabytes and Terabytes and tell us that this data contains so many distinct categories?

As an example, consider the voter's data. By considering some inputs from each voter (these are called features in AI terminology), let the machine predict that there are so many voters who would vote for X political party and so many would vote for Y, and so on. Thus, in general, we are asking the machine given a huge set of data points X, "What can you tell me about X?". Or it may be a question like "What are the five best groups we can make out of X?". Or it could be even like "What three features occur together most frequently in X?".

This is exactly the Unsupervised Learning is all about.

## Algorithms for Unsupervised Learning

Let us now discuss one of the widely used algorithms for classification in unsupervised machine learning.

### k-means clustering

The 2000 and 2004 Presidential elections in the United States were close — very close. The largest percentage of the popular vote that any candidate received was 50.7% and the lowest was 47.9%. If a percentage of the voters were to have switched sides, the outcome of the election would have been different. There are small groups of voters who, when properly appealed to, will switch sides. These groups may not be huge, but with such close races, they may be big enough to change the outcome of the election. How do you find these groups of people? How do you appeal to them with a limited budget? The answer is clustering.

Let us understand how it is done.

- 

First, you collect information on people either with or without their consent: any sort of information that might give some clue about what is important to them and what will influence how they vote.

- 

-

Then you put this information into some sort of clustering algorithm.

- 
- 

Next, for each cluster (it would be smart to choose the largest one first) you craft a message that will appeal to these voters.

- 
- 

Finally, you deliver the campaign and measure to see if it's working.

- 

Clustering is a type of unsupervised learning that automatically forms clusters of similar things. It is like automatic classification. You can cluster almost anything, and the more similar the items are in the cluster, the better the clusters are. In this chapter, we are going to study one type of clustering algorithm called k-means. It is called k-means because it finds 'k' unique clusters, and the center of each cluster is the mean of the values in that cluster.

## Cluster Identification

Cluster identification tells an algorithm, "Here's some data. Now group similar things together and tell me about those groups." The key difference from classification is that in classification you know what you are looking for. While that is not the case in clustering.

Clustering is sometimes called unsupervised classification because it produces the same result as classification does but without having predefined classes.

Now, we are comfortable with both supervised and unsupervised learning. To understand the rest of the machine learning categories, we must first understand Artificial Neural Networks (ANN), which we will learn in the next chapter.

References

- [https://www.tutorialspoint.com/machine\\_learning/](https://www.tutorialspoint.com/machine_learning/)