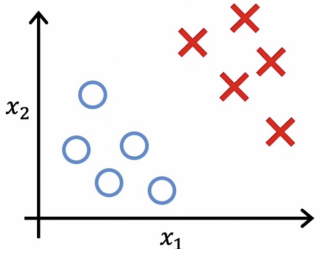
Clustering is one of the most famous unsupervised learning algorithms

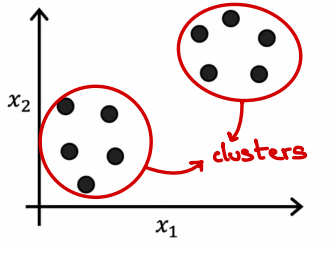
The main difference between supervised and unsupervised learning is that you are not given any correct labels (y), instead the algorithm finds relative information about the data on its own.

Supervised learning



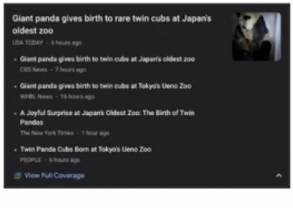
Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots, (x^{(m)}, y^{(m)})\}$

Unsupervised learning

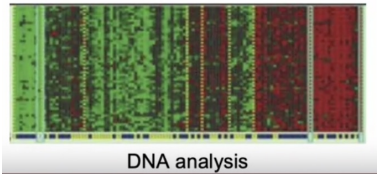


Training set: $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$

Some examples are grouping similar news together, DNA analysis, clustering group of bodies in space to find out which one is a galaxy or other coherent structures in space.



Grouping similar news



DNA analysis

Image credit: NASA/JPL-Calte
ch/E. Churchwell (Univ. of Wisconsin, Madison)

Astronomical data analysis