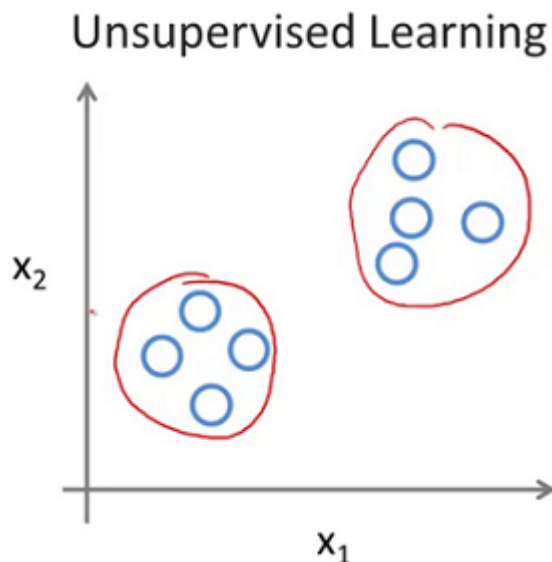


Unsupervised Learning

Unsupervised learning is a type of machine learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision.

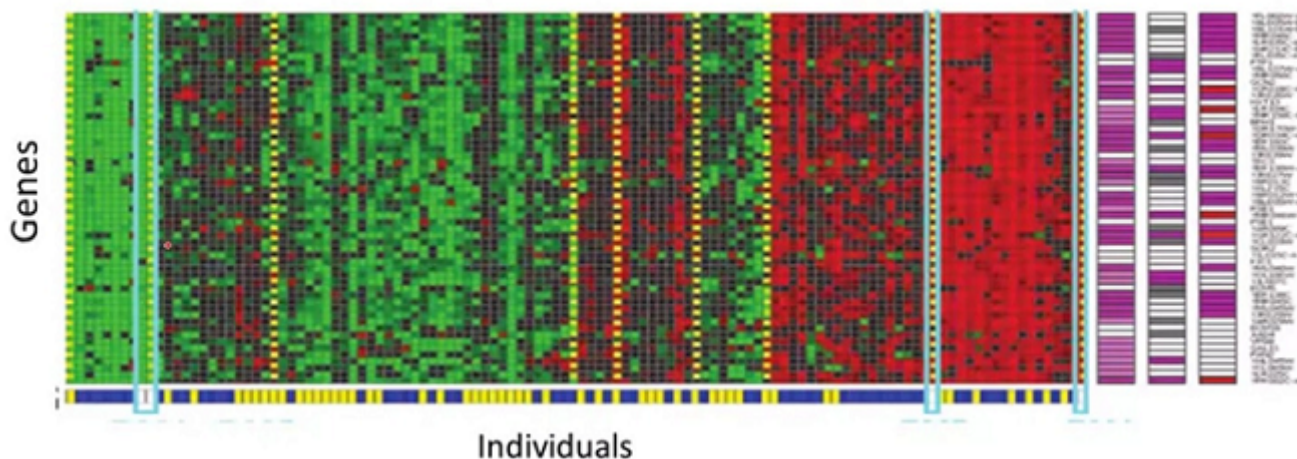
In Unsupervised Learning, we're given data that doesn't have any labels or that all has the same label or really no labels. So we're given the data set and we're not told what to do with it and we're not told what each data point is.



An Unsupervised Learning algorithm might decide that the data lives in two different clusters. Supervised Learning algorithm may break these data into these two separate clusters. So this is called a **clustering algorithm**. Two examples where clustering is used is in Google News and Understanding genomics.

An example of DNA microarray data (for understanding genomics). The idea is put a group of different individuals and for each of them, we measure how much they do or do not have a certain gene.

Technically we measure how much certain genes are expressed. So the colors, red, green, gray and so on, they show the degree to which different individuals do or do not have a specific gene.



And what we can do is then run a clustering algorithm to group individuals into different categories or into different types of people. This is Unsupervised Learning because we're not telling the algorithm in advance that who are type 1 people, who are type 2 persons, who are type 3 persons and so on and instead what we're saying is here's a bunch of data.

Since we are not giving the correct answer to the algorithm for the dataset examples, this is unsupervised learning. Unsupervised Learning or clustering is used for a bunch of other applications. It's used to organize large computer clusters.

Other unsupervised learning applications:

- Organize computing clusters
- Social network analysis
- Market segmentation
- Astronomical data analysis

Video Question: Of the following examples, which would you address using an unsupervised learning algorithm? (Check all that apply.)

- Given email labeled as spam/not spam, learn a spam filter.

Given a set of news articles found on the web, group them into sets of articles about the same stories.

Given a database of customer data, automatically discover market segments and group customers into different market segments.

- Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

Summary of Unsupervised Learning

Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.

We can derive this structure by clustering the data based on relationships among the variables in the data. With unsupervised learning there is no feedback based on the prediction results.

Example:

Clustering: Take a collection of 1,000,000 different genes, and find a way to automatically group these genes into groups that are somehow similar or related by different variables, such as lifespan, location, roles, and so on.

Non-clustering: The "Cocktail Party Algorithm", allows you to find structure in a chaotic environment. (i.e. identifying individual voices and music from a mesh of sounds at a cocktail party).