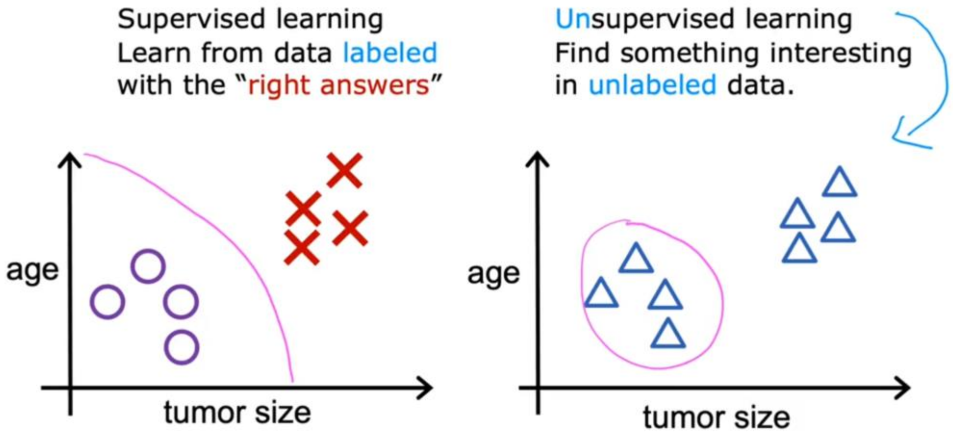
Apprentissage non supervisé partie 1

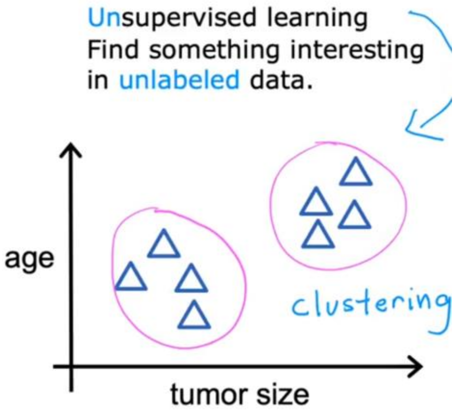
When we're looking at supervised learning in the last video recalled, it looks something like this in the case of a classification problem. Each example, was associated with an output label y such as benign or malignant, designated by the poles and crosses in unsupervised learning. Were given data that isn't associated with any output labels y, say you're given data on patients and their tumor size and the patient's age. But not whether the tumor was benign or malignant, so the dataset looks like this on the right. We're not asked to diagnose whether the tumor is benign or malignant, because we're not given any labels. Why in the dataset, instead, our job is to find some structure or some pattern or just find something interesting in the data. This is unsupervised learning, we call it unsupervised because we're not trying to supervise the algorithm. To give some quote right answer for every input, instead, we asked the our room to figure out all by yourself what's interesting.



Supervised learning
Learn from data labeled
with the "right answers"

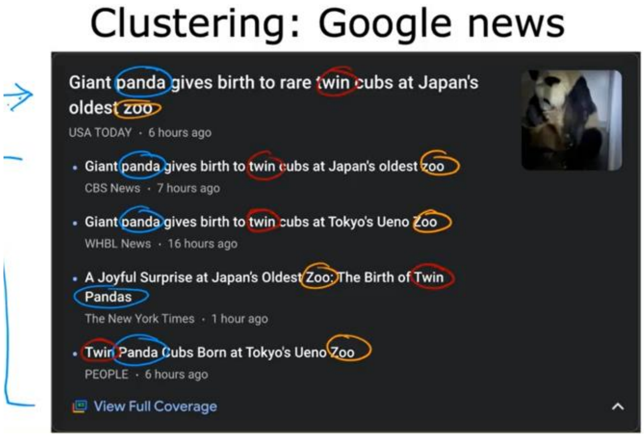
Unsupervised learning
Find something interesting
in unlabeled data.

Or what patterns or structures that might be in this data, with this particular data set. An unsupervised learning algorithm, might decide that the data can be assigned to two different groups or two different clusters. And so it might decide, that there's one cluster what group over here, and there's another cluster or group over here. This is a particular type of unsupervised learning, called a clustering algorithm. Because it places the unlabeled data, into different clusters and this turns out to be used in many applications.



Unsupervised learning
Find something interesting
in unlabeled data.

For example, clustering is used in google news, what google news does is every day it goes. And looks at hundreds of thousands of news articles on the internet, and groups related stories together. For example, here is a sample from Google News, where the headline of the top article, is giant panda gives birth to rear twin cubs at Japan's oldest zoo. Notice that the word panda appears here here, here, here and here and notice that the word twin also appears in all five articles. And the word Zoo also appears in all of these articles, so the clustering algorithm is finding articles. All of all the hundreds of thousands of news articles on the internet that day, finding the articles that mention similar words and grouping them into clusters. Now, what's cool is that this clustering algorithm figures out on his own which words suggest, that certain articles are in the same group. What I mean is there isn't an employee at google news who's telling the algorithm to find articles that the word panda. And twins and zoo to put them into the same cluster, the news topics change every day. And there are so many news stories, it just isn't feasible to people doing this every single day for all the topics that use covers. Instead the algorithm has to figure out on his own without supervision, what are the clusters of news articles today. So that's why this clustering algorithm, is a type of unsupervised learning algorithm.



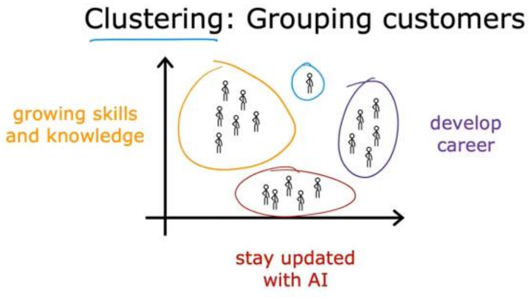
Clustering: Google news

Giant panda gives birth to rare twin cubs at Japan's oldest zoo
USA TODAY · 6 hours ago

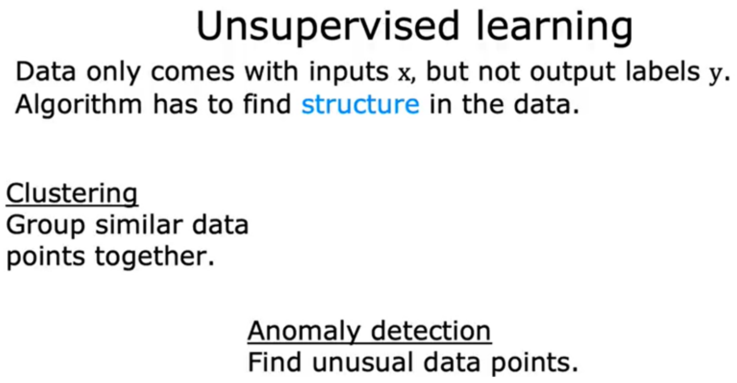
- Giant panda gives birth to twin cubs at Japan's oldest zoo
  CBS News · 7 hours ago
- Giant panda gives birth to twin cubs at Tokyo's Ueno Zoo
  WHBL News · 16 hours ago
- A Joyful Surprise at Japan's Oldest Zoo: The Birth of Twin Pandas
  The New York Times · 1 hour ago
- Twin Panda Cubs Born at Tokyo's Ueno Zoo
  PEOPLE · 6 hours ago

This is unsupervised learning, here's the third example, many companies have huge databases of customer information given this data. Can you automatically group your customers, into different market segments so that you can more efficiently serve your customers. Concretely the deep learning dot AI team did some research to better understand the deep learning dot AI community. And why different individuals take these classes, subscribed to the batch weekly newsletter, or attend our AI events. Let's visualize the deep learning dot AI community, as this collection of people running clustering. That is market segmentation found a few distinct groups of individuals, one group's primary motivation is seeking knowledge to grow their skills.

a second group's primary motivation is looking for a way to develop their career. And yet another group wants to stay updated on how AI impacts their field of work. This is a clustering that our team used to try to better serve our community as we're trying to figure out. Whether the major categories of learners in the deeper and community, So if any of these is your top motivation for learning, that's great. And I hope I'll be able to help you on your journey, or in case this is you, and you want something totally different than the other three categories. That's fine too, and I want you to know, I love you all the same, so to summarize a clustering algorithm. Which is a type of unsupervised learning algorithm, takes data without labels and tries to automatically group them into clusters.



In unsupervised learning, the data comes only with inputs x but not output labels y, and the algorithm has to find some structure or some pattern or something interesting in the data. We're seeing just one example of unsupervised learning called a clustering algorithm, which groups similar data points together. In this specialization, you'll learn about clustering as well as two other types of unsupervised learning. One is called anomaly detection, which is used to detect unusual events. This turns out to be really important for fraud detection in the financial system, where unusual events, unusual transactions could be signs of fraud and for many other applications.



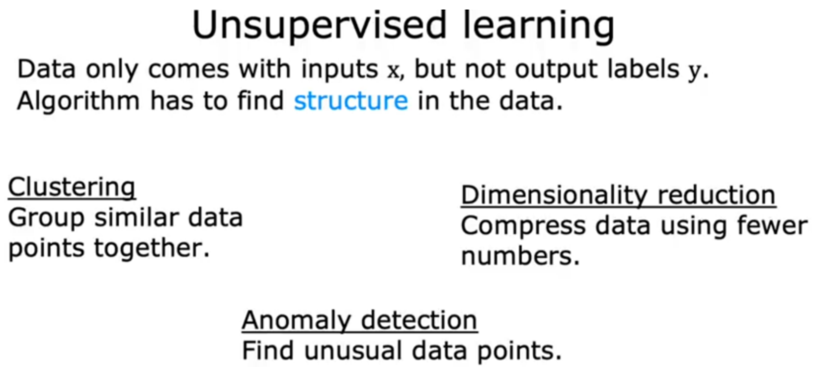
Unsupervised learning

Data only comes with inputs x, but not output labels y.
Algorithm has to find structure in the data.

Clustering
Group similar data
points together.

Anomaly detection
Find unusual data points.

You also learn about dimensionality reduction. This lets you take a big data-set and almost magically compress it to a much smaller data-set while losing as little information as possible. In case anomaly detection and dimensionality reduction don't seem to make too much sense to you yet. Don't worry about it. We'll get to this later in the specialization.



Unsupervised learning

Data only comes with inputs x, but not output labels y.
Algorithm has to find structure in the data.

Clustering
Group similar data
points together.

Dimensionality reduction
Compress data using fewer
numbers.

Anomaly detection
Find unusual data points.