Supervised vs Unsupervised machine learning

Course	
Topic	
Type	Lecture
Done	☑
Date	@July 11, 2022
Materials	https://drive.google.com/file/d/1EQC2CnTYN7EUieKc_0v7RkATU1IZGDkO/view?usp=sharing

▼ Summary

Supervised machine learning

Types of Supervised Learning: Regression and Classification

Unsupervised machine learning

Types of Unsupervised Learning: Clustering, Anomaly Detection and Dimensionality Reduction

Supervised Machine Learning

- Supervised machine learning or more commonly, supervised learning, refers to algorithms that learn x to y or input to output mappings.
- The key characteristic of supervised learning is that you give your learning algorithm examples to learn from.
- That includes the right answers, whereby right answer, I mean, the correct label y for a given input x, and is by seeing correct pairs of input x and desired output label y that the

learning algorithm eventually learns to take just the input alone without the output label and gives a reasonably accurate prediction or guess of the output.

- For example,

Input (X)	Output (Y)	Application
email	spam? (0/1)	spam filtering
audio	text transcripts	speech recognition
English	Spanish	machine translation
ad, user info	click? (0/1)	online advertising
image, radar info	position of other cars	self-driving car
image of phone	defect? (0/1)	visual inspection

Figure 1. Some examples of Supervised learning

Types of Supervised Learning:

1) Regression:

- Simple Example, trying to predict house prices based on size of the house.
- One way to solve this is to draw a straight line to the data and reading off the straight line
- But we can use other methods too. For example, we can draw a curve to fit the data.
- The task of the learning algorithm is to produce more of these right answers. So you can choose any method.
- This housing price prediction is the particular type of supervised learning called **regression**.
- The main task of regression is to predict a number.

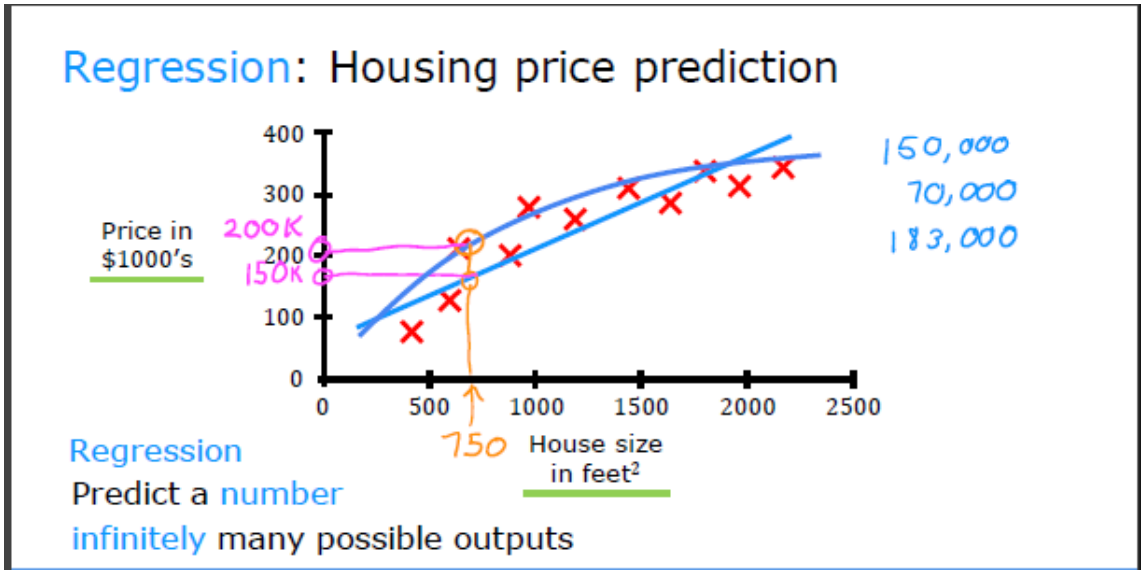

Figure 2: Regression house price prediction problem

2) Classification:

- For example, a breast cancer detection system. Tumors are usually malignant or benign.
- We can represent a tumour as this:

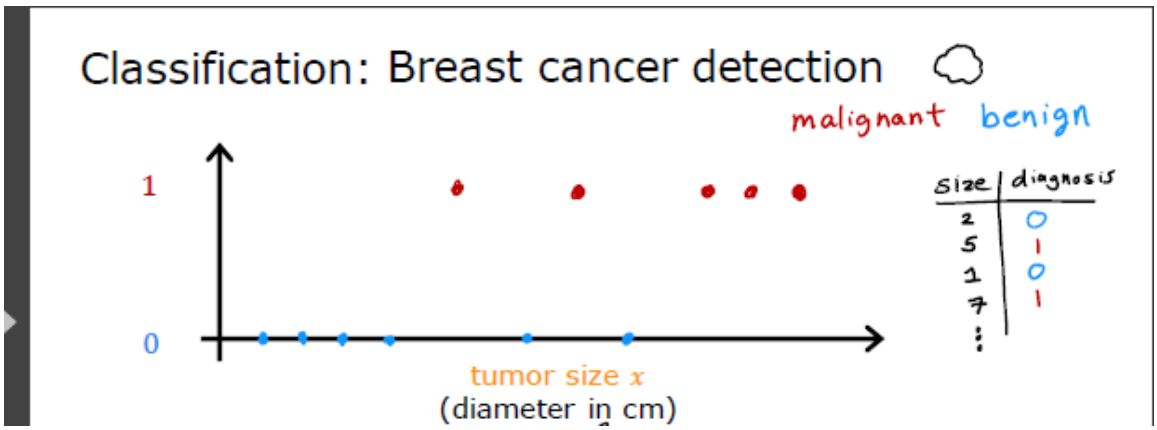

Figure 2: Representing breast cancer tumors as a classification problems. (0, benign) (1,malignant)

- One reason that this is different from regression is that we're trying to predict only a small number of possible outputs or categories. In this case two possible outputs 0 or 1, benign or malignant.

> 💡 This is different from regression which tries to predict any number, all of the infinitely many number of possible numbers. And so the fact that there are only two possible outputs is what makes this classification. Because there are only two possible outputs or two possible categories in this example, you can also plot this data set on a line like this(Figure 2).

- Output categories(benign,malignant) is same as Output classes.
- ***So to summarize classification algorithms predict categories.*** Categories don't have to be numbers. It could be non numeric for example, it can predict whether a picture is that of a cat or a dog. And it can predict if a tumor is benign or malignant. Categories can also be numbers like 0, 1 or 0, 1, 2. But what makes classification different from regression when you're interpreting the numbers is that classification predicts a small finite limited set of possible output categories such as 0, 1 and 2 but not all possible numbers in between like 0.5 or 1.7.

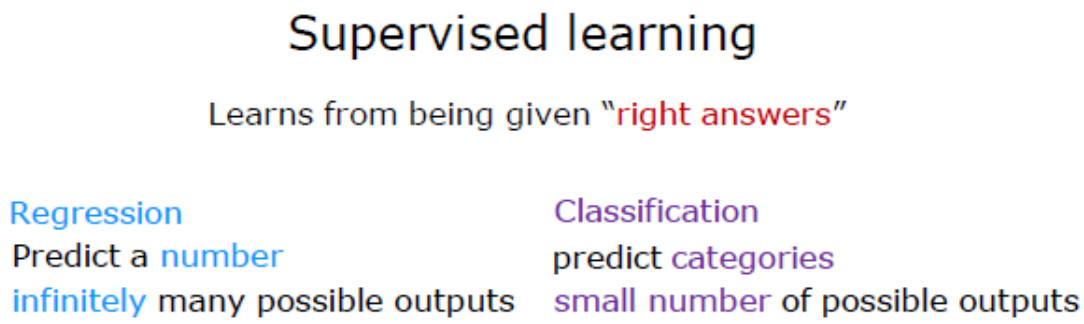

Figure 3: Summary of Supervised Learning.

Unsupervised Learning:

- We are not given labels. We are not trying to supervise the algorithm into saying what is right.

- 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.
- 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. Or what patterns or structures that might be in this data, with this particular data set.

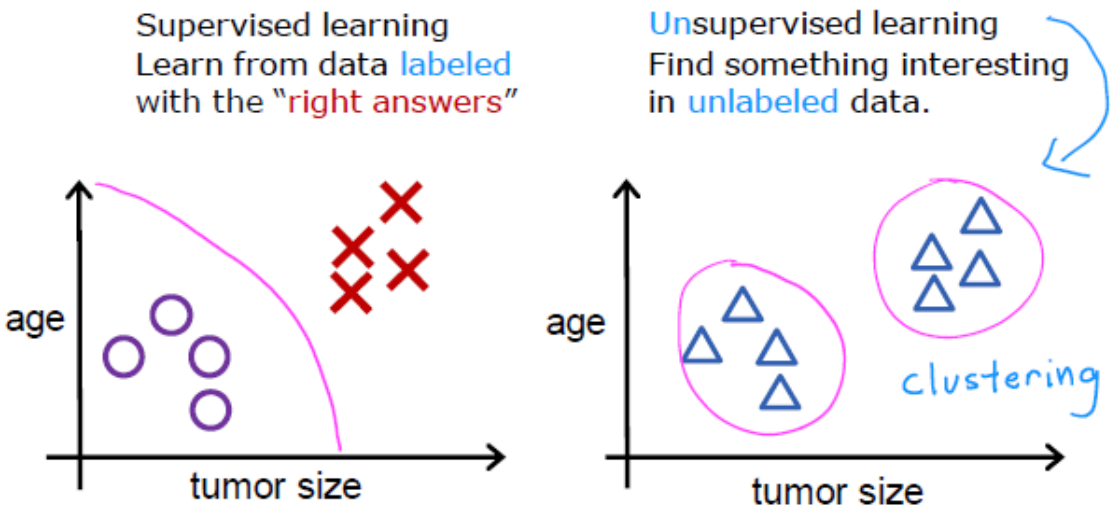

Figure 4: Unsupervised Learning example

Types of unsupervised learning:

1) Clustering algorithms:

- It is a type of unsupervised learning algorithm, takes data without labels and tries to automatically group them into clusters. Data only comes with inputs x, but not output labels y. Algorithm has to find structure in the data
- For example, trying to group google news into different clusters (Check figure 5)

Figure 5: Grouping different panda twins news via a clustering algorithm

- Can be used to identify different customer segments for a product.

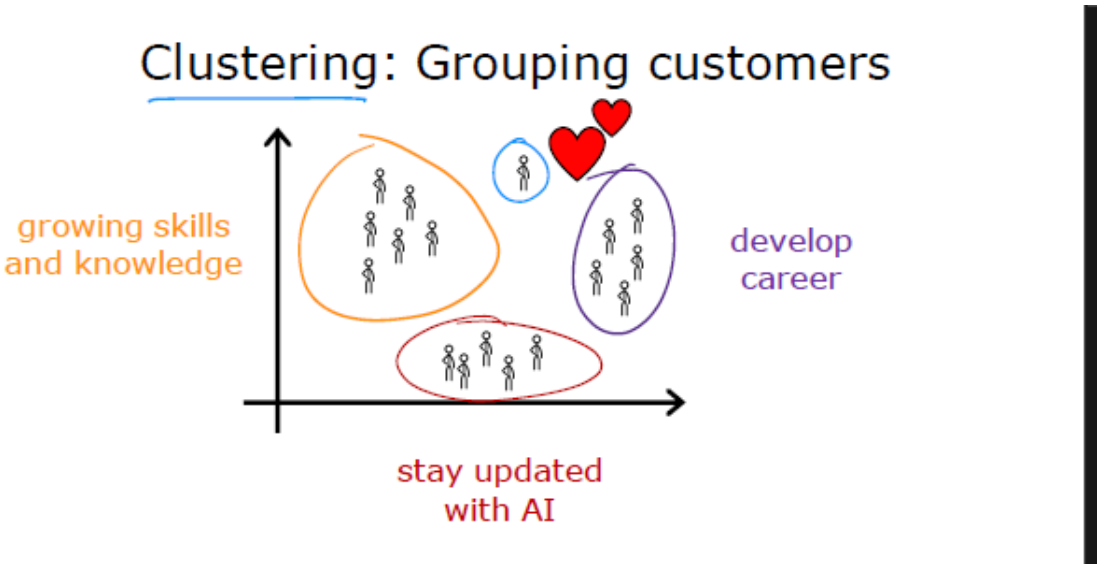

Figure 6: Grouping different customers according to their different needs of learning machine learning.

2) Anomaly Detection:

- Find unusual data points.
- For example, finding unusual credit card transactions and blocking them.

3) Dimensionality reduction:

- It 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.
- Compress data using fewer numbers.

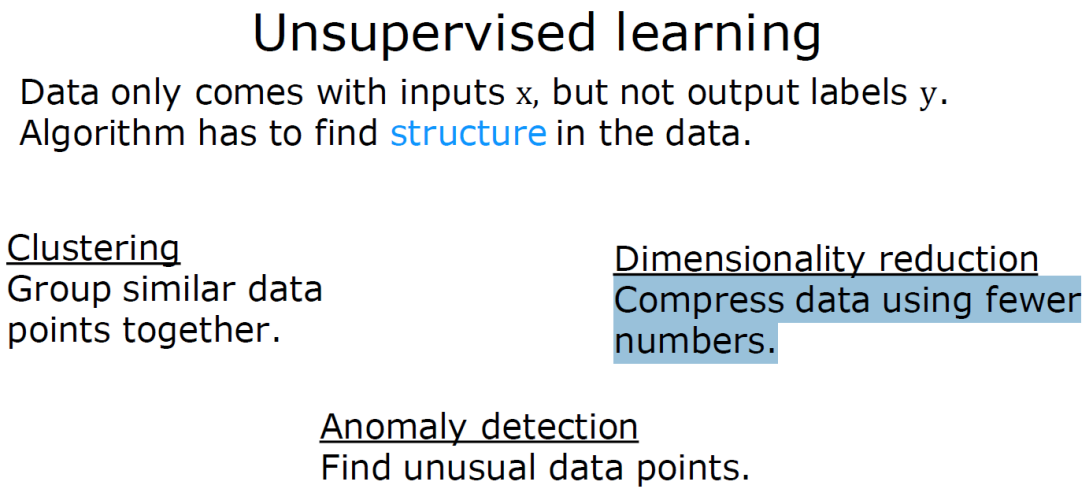

Figure 7: Summary of Unsupervised Learning.