**Supervised vs Unsupervised (5:59) 23/09**

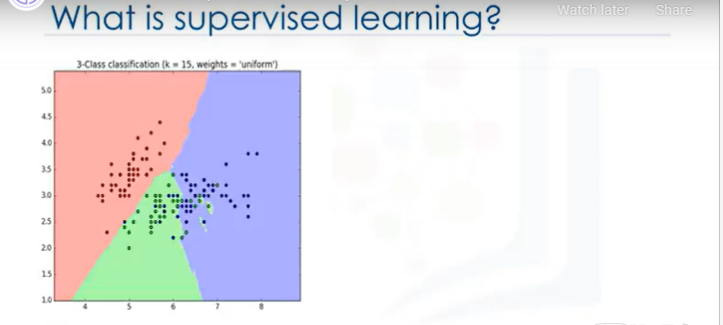
**In this video, we’ll introduce Supervised algorithms versus UnSupervised algorithms.**

**An easy way to begin grasping the concept of supervised learning is by looking directly at the words that make it up.**

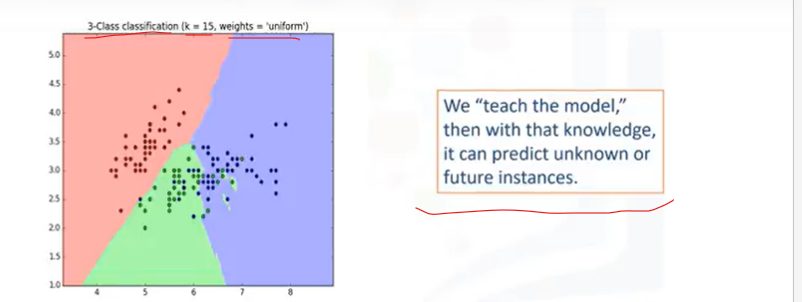
**Supervise means to observe and direct the execution of a task, project, or activity.**

Obviously, we aren’t going to be supervising a person…

1. Instead, we’ll be supervising a machine learning model that might be able to produce classification regions like we see here.



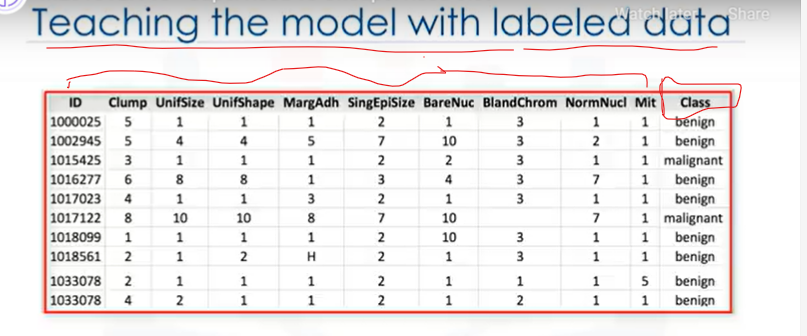
1. **So, how do we supervise a machine learning model?**
   * We do this by “teaching” the model.
   * **That is, we load the model with knowledge so that we can have it predict future instances.**
   * But … this leads to the next question
2. **How exactly do we teach a model?**
   * We teach the model by training it with some data from a labeled dataset.



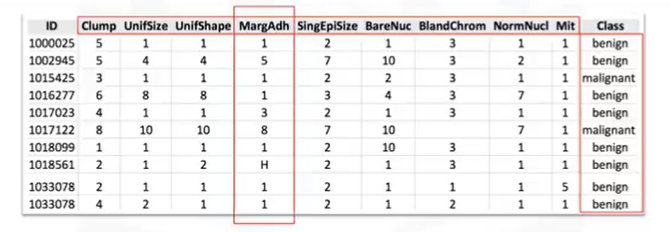
* + It’s important to note that the data is labeled.

1. And what does a labeled dataset look like?

Well, it can look something like this.



1. This example is taken from the cancer dataset.
2. As you can see, we have some historical data for patients, and we already know the class of each row.
3. Let’s start by introducing some components of this table.
4. The names up here, which are called Clump thickness, Uniformity of cell size, Uniformity of cell shape, Marginal adhesion, and so on, are called **Attributes**.
5. The columns are called **Features** (i.e X), which include the data.



1. X -> [**Independent** variables Or **Predictors** or Features or Regressors (in core stats!) or Attributes], which include the data and the
2. Y (Class) is -> [**Dependent** variable or Target variable or O/P var or Var of Interest].

The dependent variables refer to that type of variable that measures the affect of the Independent variable(s) on the test units.

The dependent variables are named as ‘Predicted Variable’ because they are the values that are predicted or assumed by the predictor / independent variables.

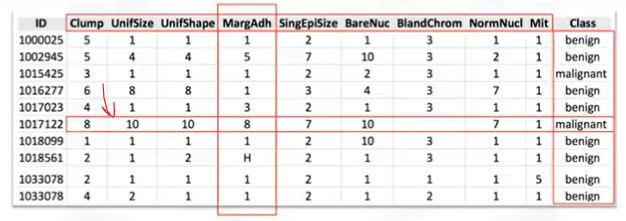
**ADDITIONALLY:**

**Independent** variables are also called “**Predictors,**” “Regressor”, “**Explanatory** variable,” “Controlled variable,” “Manipulated variable,” “Exposure variable,” and/or “Input variable.”

Similarly,

**Dependent** variables are also called “**Response** variable,” “Observed variable,” “Explained variable,” “**Outcome** variable,” “Regressand,” “Measured variable,” “Responding variable,” Experimental variable,” and/or “Output variable.”

If you plot this data, and look at a single data point on a plot, it’ll have all of **these attributes.**

**That would make a row on this chart, also referred to as an observation.**

**Looking directly at the value of the data, you can have two kinds.**

1. **The first is numerical.**

When dealing with machine learning, the most commonly used data is numeric.

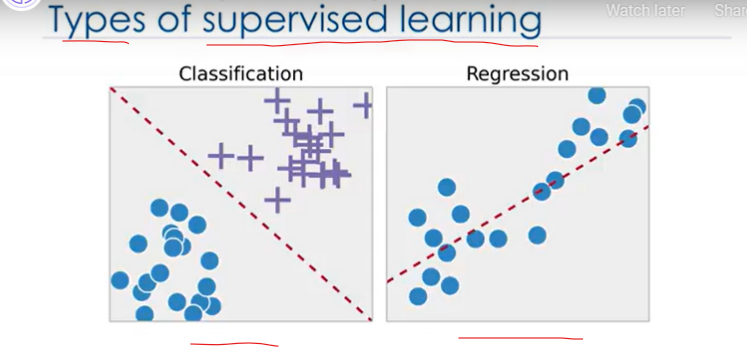
1. **The second is categorical… that is, it’s non-numeric, because it contains characters** rather than numbers.

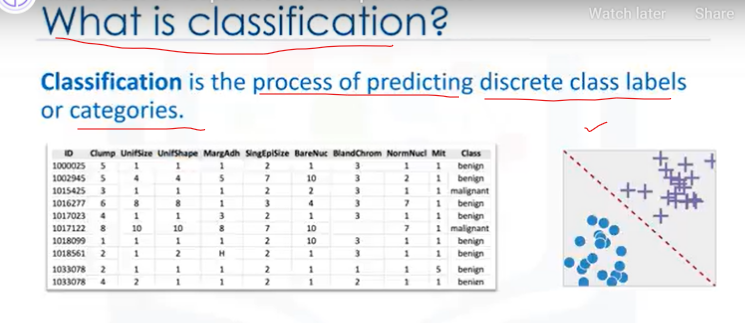


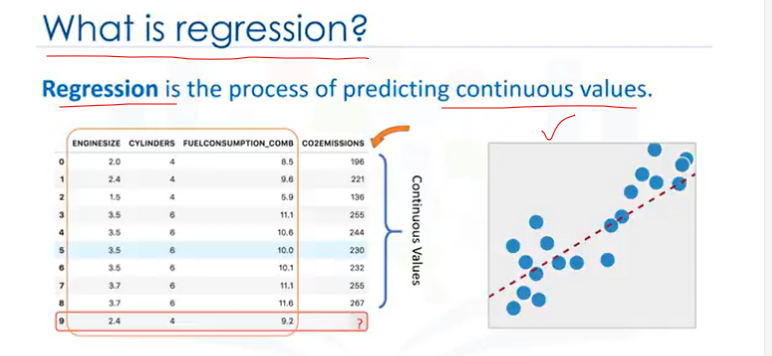
**In this case, it’s categorical because this dataset is made for Classification.**

**There are two types of Supervised Learning techniques.**

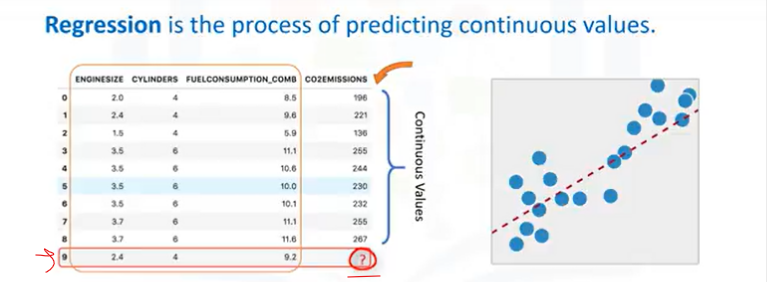
1. **They are: Classification and Regression.**
2. **Classification is the process of predicting a discrete class label or category.**
3. **Regression is the process of predicting a continuous value as opposed to predicting a categorical value in Classification.**







1. **Look at this dataset.**



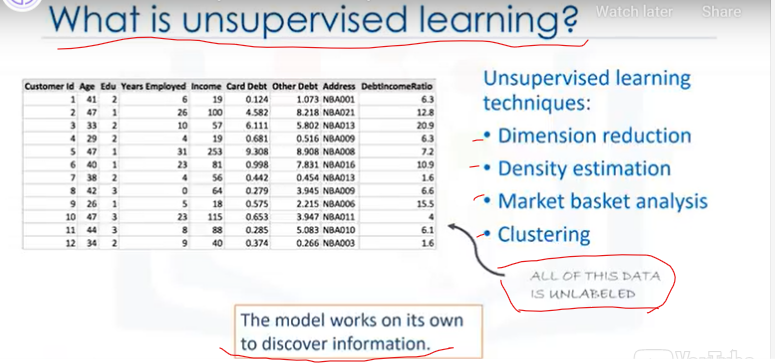
1. It is related to Co2 emissions of different cars.
2. It includes Engine size, Cylinders, Fuel Consumption and Co2 emission of various models of automobiles.
3. **Given this dataset, you can use regression to predict the Co2 emission of a new car by**
4. using other fields, such as Engine size or number of Cylinders.

Since we know the meaning of supervised learning,

**Que) What do you think unsupervised learning means?**

1. **Unsupervised Learning is exactly as it sounds.**
2. We do not supervise the model, but we let the model work on its own to discover information that may not be visible to the human eye.
3. It means, The Unsupervised algorithm trains on the dataset, and draws conclusions on UNLABELED data.

Generally speaking, unsupervised learning has more difficult algorithms than supervised learning, since we know little to no information about the data, or the outcomes that are to be expected.



**Dimension reduction, Density estimation, Market basket analysis and Clustering are the most widely used unsupervised machine learning techniques.**

**Que) What is Dimensionality Reduction (or Feature Selection)?**

1. Dimensionality Reduction and/or feature selection play a large role in this by reducing redundant **features to make the classification easier.**

**Que) What is Market Basket Analysis?**

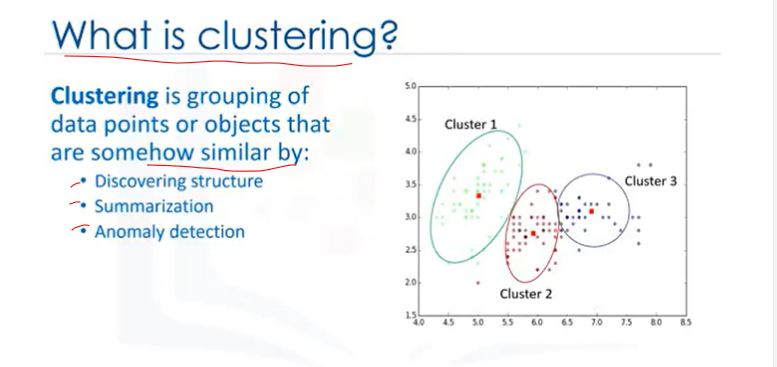
1. Market basket analysis is a modelling technique based upon the theory that if you buy a certain group of items, you’re more likely to buy another group of items.

**Que) What is Density Estimation?**

1. Density estimation is a very simple concept that is mostly used to explore the data to find some structure within it.

**Que) What is Clustering?**

1. Clustering is considered to be one of the most popular unsupervised machine learning techniques used for grouping data points or objects that are somehow similar.
2. Cluster analysis has many applications in different domains, whether it be a bank’s desire to segment its customers based on certain characteristics, or helping an individual to organize and group his/her favorite types of music!
3. Generally speaking, though, Clustering is used mostly for: Discovering structure, Summarization, and Anomaly detection.



**So, to recap, the biggest difference between Supervised and Unsupervised Learning is that!**

**See**

**Below!**

**Que) What is, the biggest difference between Supervised and Unsupervised Learning?**

1. Supervised learning deals with labeled data while Unsupervised Learning deals with unlabeled data.
2. In Supervised learning, we have machine learning algorithms for Classification and Regression.
3. In Unsupervised learning, we have methods such as clustering.
4. In comparison to Supervised learning, unsupervised learning has fewer models and fewer evaluation methods that can be used to ensure that the outcome of the model is accurate.
5. As such, Unsupervised learning creates a less controllable environment, as the machine is creating outcomes for us.



**Summary**

**What have we learned? We have learned —**

1. X -> [Independent variables Or Predictors or Features or Regressors (in core stats!) or Attributes], which include the data and the
2. Y (Class) is -> [Dependent variable or Target variable or O/P var or Var of Interest].

**ADDITIONALLY:**

1. Independent variables are also called “Regressor”, “Explanatory variable,” “Controlled variable,” “Manipulated variable,” “Exposure variable,” and/or “Input variable.”
2. Similarly,
3. Dependent variables are also called “Response variable,” “Observed variable,” “Explained variable,” “Outcome variable,” “Regressand,” “Measured variable,” “Responding variable,” Experimental variable,” and/or “Output variable.”

**There are two types of Supervised Learning techniques.**

1. **They are: Classification and Regression.**
2. **Classification is the process of predicting a discrete class label or category.**
3. **Regression is the process of predicting a continuous value as opposed to predicting a categorical value in Classification.**

**Que) What do you think unsupervised learning means?**

1. **Unsupervised Learning is exactly as it sounds.**
2. We do not supervise the model, but we let the model work on its own to discover information that may not be visible to the human eye.
3. It means, The Unsupervised algorithm trains on the dataset, and draws conclusions on UNLABELED data.

**the most widely used unsupervised machine learning techniques.**

**Que) What is Dimensionality Reduction (or Feature Selection)?**

Dimensionality Reduction and/or feature selection play a large role in this by reducing redundant **features to make the classification easier.**

**Que) What is Market Basket Analysis?**

Market basket analysis is a modelling technique based upon the theory that if you buy a certain group of items, you’re more likely to buy another group of items.

**Que) What is Density Estimation?**

Density estimation is a very simple concept that is mostly used to explore the data to find some structure within it.

**Que) What is Clustering?**

Clustering is considered to be one of the most popular unsupervised machine learning techniques used for grouping data points or objects that are somehow similar.

Cluster analysis has many applications in different domains, whether it be a bank’s desire to segment its customers based on certain characteristics, or helping an individual to organize and group his/her favorite types of music!

**Que) What is, the biggest difference between Supervised and Unsupervised Learning?**



## Difference Between Classification and Regression?

The main difference between **classification and regression** is that classification predicts or classify a discrete label such as True or False, Spam or Not Spam, etc. while regression predicts a **continuous quantity** or value such as price, salary, age, etc.

A regression algorithm is commonly evaluated by calculating the **Root Mean Squared Error?** of its output

While a Classification algorithm is evaluated by computing the Accuracy with which it correctly classified its input.

**TEST!**

### **Question 1**

**Machine Learning uses algorithms that can learn from data without relying on explicitly programmed methods.**

**True**

False

### **Question 2**

**Which are the two types of Supervised learning techniques?**

Classification and Clustering

Classification and K-Means

Regression and Clustering

Regression and Partitioning

**Classification and Regression**

### **Question 3**

**Which of the following statements best describes the Python scikit library?**

A library for scientific and high-performance computation.

**A collection of algorithms and tools for machine learning.**

A popular plotting package that provides 2D plotting as well as 3D plotting.

A library that provides high-performance, easy to use data structures.

A collection of numerical algorithms and domain-specific toolboxes.