**Relationship Between Artificial Intelligence and Machine Learning : -**

Ai refers to the broad capabilities of machines to perform activities using human like intelligence.

ML is a type of AI , it allows comps to automatically learn from experience without being explicitly programmed to do so. Using ML computers can learn from data to discover patterns and make predictions.

**There are 3 kinds of ML:-**

1. **Supervised Learning :-** Every training sample from the dataset has a corresponding label or output value associated with it. As a result , the Algo learns to predict labels or output values. Ex – predict a sale price of a house or classify objects in an image.
2. **Unsupervised Learning :** There are no labels for the training data. A ML Algo tries to learn underlying patterns or distributions that govern the data.

**In Supervised and unsupervised Learning the ML model inspects the data and tries to discover patterns, then the programmer or the user use the discovered patterns by the ML model to gain new understandings or predict.**

1. **Reinforcement Learning :-** It is a type of learning which takes a different approach. It learns through consequences of actions in an environment i.e it finds out what actions to take in certain situations to maximize the reward (somewhat like the Greedy approach of Algorithms).

**How Machine Learning differs from traditional programming based approaches –**

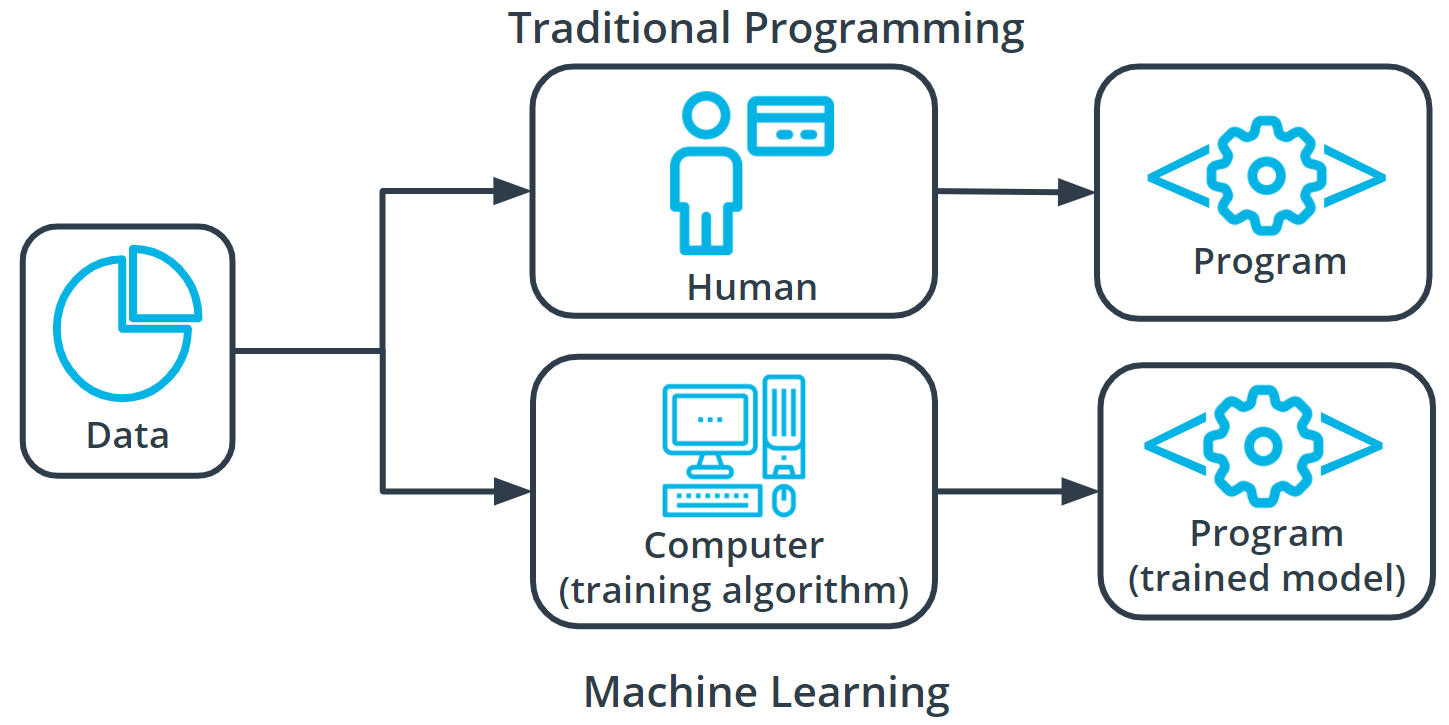
In traditional problem-solving with software, a person analyses a problem and engineers a solution in code to solve that problem. For many real-world problems, this process can be laborious (or even impossible) because a correct solution would need to consider a vast number of edge cases.

Imagine, for example, the challenging task of writing a program that can detect if a cat is present in an image. Solving this in the traditional way would require careful attention to details like varying lighting conditions, different types of cats, and various poses a cat might be in.

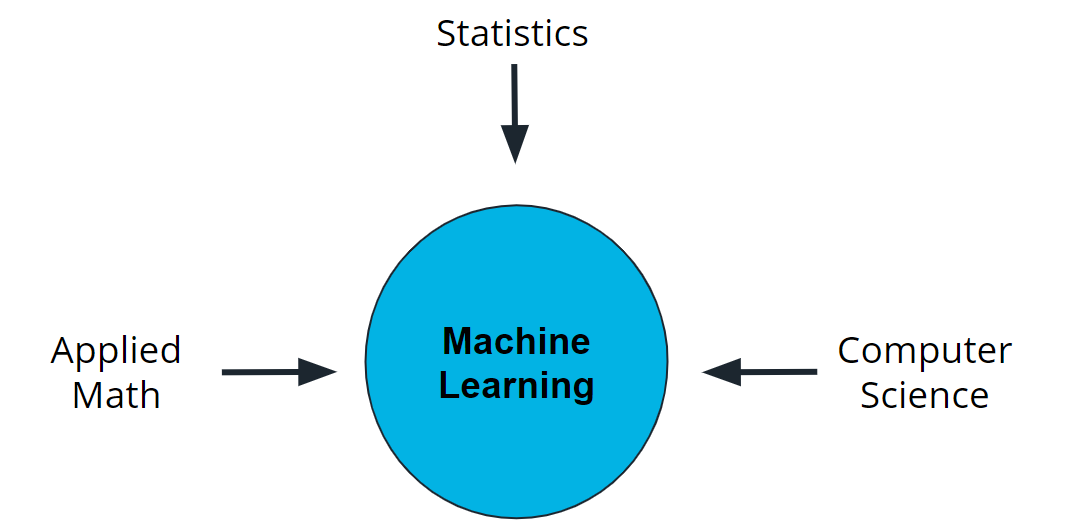
In machine learning, the problem solver abstracts away part of their solution as a flexible component called a model, and uses a special program called a model training algorithm to adjust that model to real-world data. The result is a trained model which can be used to predict outcomes that are not part of the data set used to train it.

In a way, machine learning automates some of the statistical reasoning and pattern-matching the problem solver would traditionally do.

The overall goal is to use a model created by a model training algorithm to generate predictions or find patterns in data that can be used to solve a problem.



Ex – If we train our pet dog to do certain things and reward him every time he does that correctly he will always do that because he knows that doing that stuff will reward him that’s how reinforcement learning works. It always find the best solution possible which will give maximum reward to the user.

**Understanding Terminology** :-

Machine learning is a new field created at the intersection of statistics, applied math, and computer science. Because of the rapid and recent growth of machine learning, each of these fields might use slightly different formal definitions of the same terms.

**Terminology**

**Machine learning, or ML**, is a modern software development technique that enables computers to solve problems by using examples of real-world data

**In supervised learning**, every training sample from the dataset has a corresponding label or output value associated with it. As a result, the algorithm learns to predict labels or output values.

**In reinforcement learning**, the algorithm figures out which actions to take in a situation to maximize a reward (in the form of a number) on the way to reaching a specific goal.

**In unsupervised learning**, there are no labels for the training data. A machine learning algorithm tries to learn the underlying patterns or distributions that govern the data.

**Components of Machine Learning :-**

Nearly all solvable task of Machine Learning consists of 3 steps-

1 – **Machine Learning Model**

**2 – A Machine Learning Algorithm**

**3 – A model interference Algorithm**

**Ex – Let the task be making a Clay Teapot from scratch-**

First , we will start with a block of raw clay. At this stage, the raw clay can be molded into different forms and be used to serve many different purposes. Here this is the **Machine Learning Model.**

Then , I inspect and analyse based on the clay that how I should approach to make it look like a teapot. This stage is similar to **Model Training Algorithm.**

Last but not least , Now you have our completed teapot. You inspected the clay, evaluated the changes that needed to be made, and made them, and now the teapot is ready for you to use. Enjoy your tea!

So what does this mean from a machine learning perspective? We are ready to use the model inference algorithm to generate predictions using the trained model. This process is often referred to as **Model inference Algorithm.**

**What are machine learning models?**

A machine learning model, like a piece of clay, can be molded into many different forms and serve many different purposes. A more technical definition would be that a machine learning model is a block of code or framework that can be modified to solve different but related problems based on the data provided.

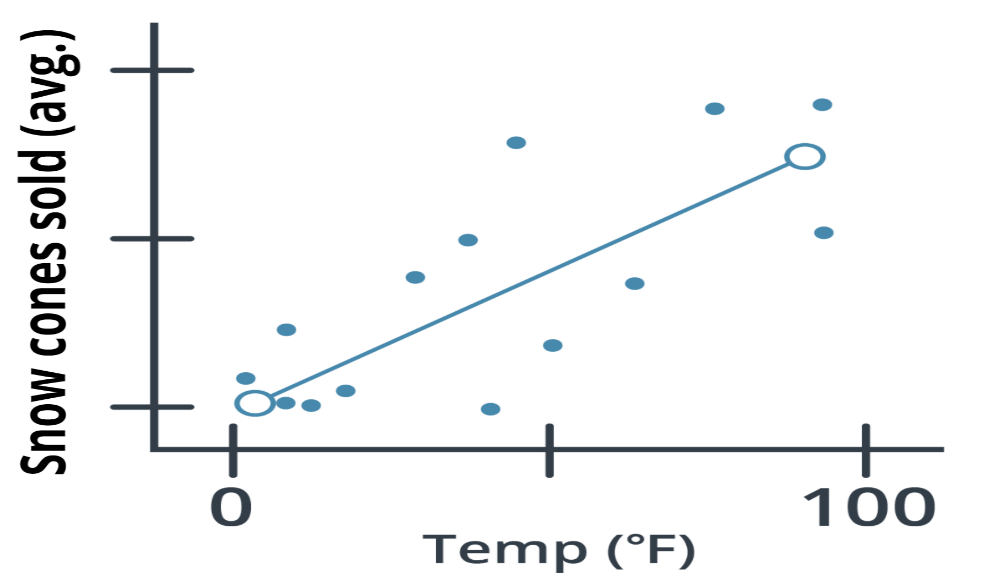
**Important**

*A model is an extremely generic program(or block of code), made specific by the data used to train it. It is used to solve different problems.*

**Two simple examples**

**Example 1**

Imagine you own a snow cone cart, and you have some data about the average number of snow cones sold per day based on the high temperature. You want to better understand this relationship to make sure you have enough inventory on hand for those high sales days.



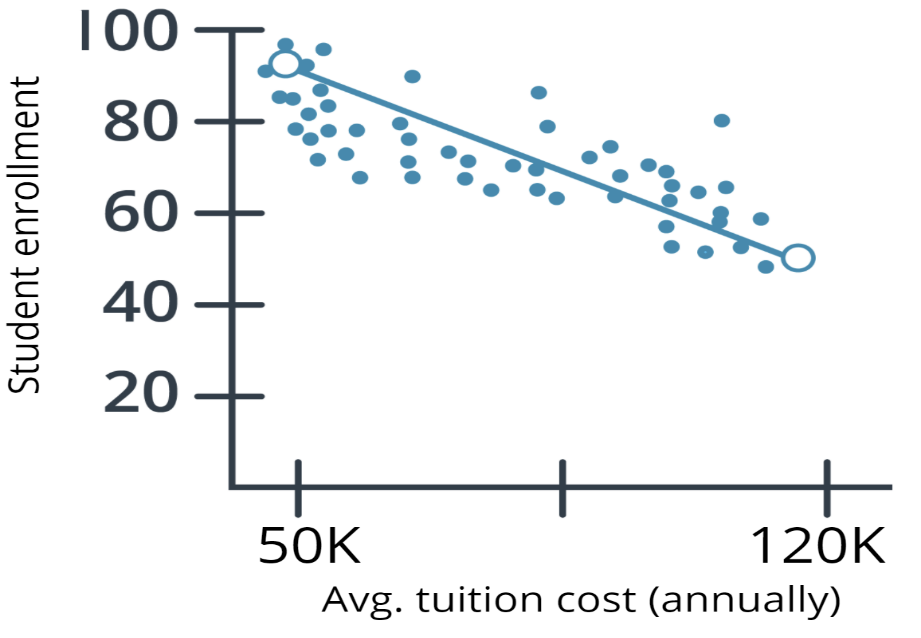
Snow cones sold regression chart

In the graph above, you can see one example of a model, a *linear regression model* (indicated by the solid line). You can see that, based on the data provided, the model predicts that as the high temperate for the day *increases* so do the average number of snow cones sold. Sweet!

Example 2

Let's look at a different example that uses the same *linear regression model*, but with different data and to answer completely different questions.

Imagine that you work in higher education and you want to better understand the relationship between the cost of enrollment and the number of students attending college. In this example, our model predicts that as the cost of tuition increases the number of people attending college is likely to decrease.



Average tuition regression chart

Using the same linear regression model (indicated by the solid line), you can see that the number of people attending college does go down as the cost increases.

*Both examples showcase that a model is a generic program made specific by the data used to train it.*

**Model Training**

How are model training algorithms used to train a model?

In the preceding section, we talked about two key pieces of information: a model and data. In this section, we show you how those two pieces of information are used to create a trained model. This process is called *model training*.

**Model training algorithms work through an interactive process**

Let's revisit our clay teapot analogy. We've gotten our piece of clay, and now we want to make a teapot. Let's look at the algorithm for molding clay and how it resembles a machine learning algorithm.

**Think about the changes that need to be made.** The first thing you would do is

inspect the raw clay and think about what changes can be made to make it look more like a teapot. Similarly, a model training algorithm uses the model to process data and then compares the results against some end goal, such as our clay teapot.

**Make those changes**. Now, you mold the clay to make it look more like a teapot. Similarly, a model training algorithm gently nudges specific parts of the model in a direction that brings the model closer to achieving the goal.

**Repeat.** By iterating over these steps over and over, you get closer and closer to what you want until you determine that you’re close enough that you can stop.

|  |  |
| --- | --- |
| clay analogy for machine learning Think about the changes that need to be made | Molding clay analoogy for machine learning Make those changes |

**Model Inference: Using Your Trained Model**

Now you have our completed teapot. You inspected the clay, evaluated the changes that needed to be made, and made them, and now the teapot is ready for you to use. Enjoy your tea!

*So what does this mean from a machine learning perspective?* We are ready to use the model inference algorithm to generate predictions using the trained model. This process is often referred to as model inference.

**5 Major Steps in Machine Learning Process.**

In the preceding diagram, you can see an outline of the major steps of the machine learning process. Regardless of the specific model or training algorithm used, machine learning practitioners practice a common workflow to accomplish machine learning tasks.

These steps are iterative. In practice, that means that at each step along the process, you review how the process is going. Are things operating as you expected? If not, go back and revisit your current step or previous steps to try and identify the breakdown.

**STEP -1 : DEFINING A PROBLEM (OUT OF 5 MAJOR STEPS) :-**

How do You Start a Machine Learning Task?

* Define a very specific task.
* Think back to the snow cone sales example. Now imagine that you own a frozen treats store and you sell snow cones along with many other products. You wonder, "‘How do I increase sales?" It's a valid question, but it's the opposite of a very specific task. The following examples demonstrate how a machine learning practitioner might attempt to answer that question.
* “Does adding a $1.00 charge for sprinkles on a hot fudge sundae increase the sales of hot fudge sundaes?”
* “Does adding a $0.50 charge for organic flavors in your snow cone increase the sales of snow cones?”

**Identify the machine learning task we might use to solve this problem.**

This helps you better understand the data you need for a project.

**What is a Machine Learning Task?**

All model training algorithms, and the models themselves, take data as their input. Their outputs can be very different and are classified into a few different groups based on the task they are designed to solve. Often, we use the kind of data required to train a model as part of defining a machine learning task.

In this lesson, we will focus on two common machine learning tasks:

1. **Supervised Learning.**
2. **Unsupervised Learning.**
3. **Reinforcement Learning**
4. **Supervised Learning or Labelled data :-** In this kind of learning we use a dataset which has some particular lables associated with them. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labelled data.

For instance, suppose you are given a basket filled with different kinds of fruits. Now the first step is to train the machine with all different fruits one by one like this:

If the shape of the object is rounded and has a depression at the top, is red in color, then it will be labeled as –**Apple**.

If the shape of the object is a long curving cylinder having Green-Yellow color, then it will be labeled as –**Banana**.

Now suppose after training the data, you have given a new separate fruit, say Banana from the basket, and asked to identify it.

Since the machine has already learned the things from previous data and this time have to use it wisely. It will first classify the fruit with its shape and color and would confirm the fruit name as **BANANA** and put it in the Banana category. Thus the machine learns the things from training data(basket containing fruits) and then applies the knowledge to test data(new fruit).

**Supervised learning classified into two categories of algorithms**:

**Classification:** A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.

**Regression:** A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Supervised learning deals with or learns with “labelled” data. This implies that some data is already tagged with the correct answer.

**Types:-**

**Regression**

**Logistic Regression**

**Classification**

**Naive Bayes Classifiers**

**K-NN (k nearest neighbors)**

**Decision Trees**

**Support Vector Machine**

**Advantages:-**

* Supervised learning allows collecting data and produces data output from previous experiences.
* Helps to optimize performance criteria with the help of experience.
* Supervised machine learning helps to solve various types of real-world computation problems.

**Disadvantages:-**

* Classifying big data can be challenging.
* Training for supervised learning needs a lot of computation time. So, it requires a lot of time.

**2 - Unsupervised learning or Unlabelled Data -**

Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself.

For instance, suppose it is given an image having both dogs and cats which it has never seen.

Thus the machine has no idea about the features of dogs and cats so we can’t categorize it as ‘dogs and cats ‘. But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts. The first may contain all pics having dogs in it and the second part may contain all pics having cats in it. Here you didn’t learn anything before, which means no training data or examples.

It allows the model to work on its own to discover patterns and information that was previously undetected. It mainly deals with unlabelled data.

**Unsupervised learning is classified into two categories of algorithms:**

**Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

**Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

**Types of Unsupervised Learning:-**

**Clustering**

**Exclusive (partitioning)**

**Agglomerative**

**Overlapping**

**Probabilistic**

**Clustering Types:-**

**Hierarchical clustering**

**K-means clustering**

**Principal Component Analysis**

**Singular Value Decomposition**

**Independent Component Analysis**

**Supervised vs. Unsupervised Machine Learning**

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Supervised machine learning** | **Unsupervised machine learning** |
| **Input Data** | Algorithms are trained using labeled data. | Algorithms are used against data that is not labeled |
| **Computational Complexity** | Simpler method | Computationally complex |
| **Accuracy** | Highly accurate | Less accurate |

**How do we classify tasks when we don't have a label?**

Unsupervised learning involves using data that doesn't have a label. One common task is called **clustering**. Clustering helps to determine if there are any naturally occurring groupings in the data.

Let's look at an example of how clustering in unlabeled data works.

**Identifying book micro-genres with unsupervised learning**

Imagine that you work for a company that recommends books to readers.

The assumption: You are fairly confident that micro-genres exist, and that there is one called Teen Vampire Romance. Because you don’t know which micro-genres exist, you can't use **supervised learning** techniques.

This is where the **unsupervised learning** clustering technique might be able to detect some groupings in the data. The words and phrases used in the book description might provide some guidance on a book's micro-genre.

In **supervised**learning, there are two main identifiers you will see in machine learning:

* A **categorical**label*has a*discrete*set of possible values. In a machine learning problem in which you want to identify the type of flower based on a picture, you would train your model using images that have been labeled with the categories of flower you would want to identify. Furthermore, when you work with categorical labels, you often carry out*classification tasks\*, which are part of the supervised learning family.
* A **continuous**(regression) label*does not have a discrete set of possible values, which often means you are working with numerical data. In the snow cone sales example, we are trying to predict the*number\* of snow cones sold. Here, our label is a number that could, in theory, be any value.

In unsupervised learning, **clustering**is just one example. There are many other options, such as deep learning.

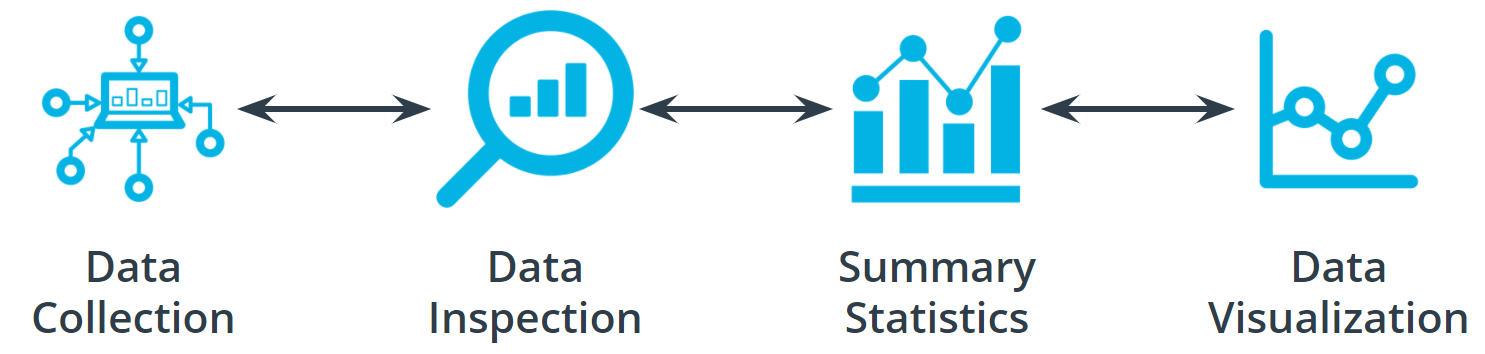
**Terminology :-**

* **Clustering.** Unsupervised learning task that helps to determine if there are any naturally occurring groupings in the data.
* A ***categorical label*** has a discrete set of possible values, such as "is a cat" and "is not a cat."
* ***A continuous (regression) label*** does not have a discrete set of possible values, which means possibly an unlimited number of possibilities.
* **Discrete:** A term taken from statistics referring to an outcome taking on only a finite number of values (such as days of the week).
* A **label** refers to data that already contains the solution.
* Using **unlabeled** data means you don't need to provide the model with any kind of label or solution while the model is being trained.

**STEP – 2 : BUILD A DATASET (OUT OF 5 MAJOR STEPS):-**

The next step in the machine learning process is to build a dataset that can be used to solve your machine learning-based problem. Understanding the data needed helps you select better models and algorithms so you can build more effective solutions.

**The most important step of the machine learning process**

Working with data is perhaps the most overlooked—yet most important—step of the machine learning process. In 2017, an O’Reilly study showed that machine learning practitioners spend 80% of their time working with their data.

You can take an entire class just on working with, understanding, and processing data for machine learning applications. Good, high-quality data is essential for any kind of machine learning project. Let's explore some of the common aspects of working with data.

**Data collection :-**

Data collection can be as straightforward as running the appropriate SQL queries or as complicated as building custom web scraper applications to collect data for your project. You might even have to run a model over your data to generate needed labels. Here is the fundamental question:

*Does the data you've collected match the machine learning task and problem you have defined?*

**Data inspection :-**

The quality of your data will ultimately be the largest factor that affects how well you can expect your model to perform. As you inspect your data, look for

* Outliers
* Missing or incomplete values
* Data that needs to be transformed or preprocessed so it's in the correct format to be used by your model

**Summary statistics :-**

Models can assume how your data is structured.

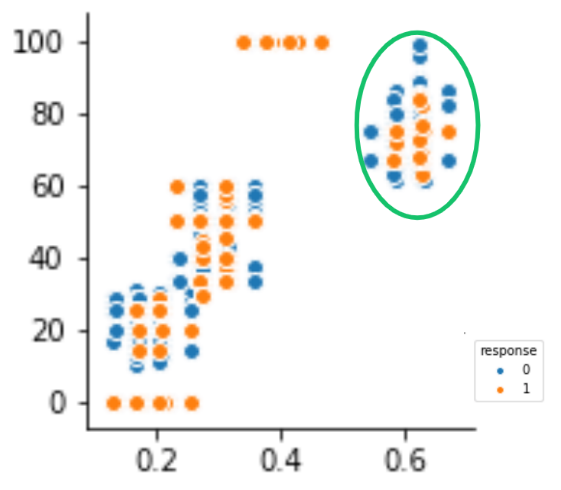
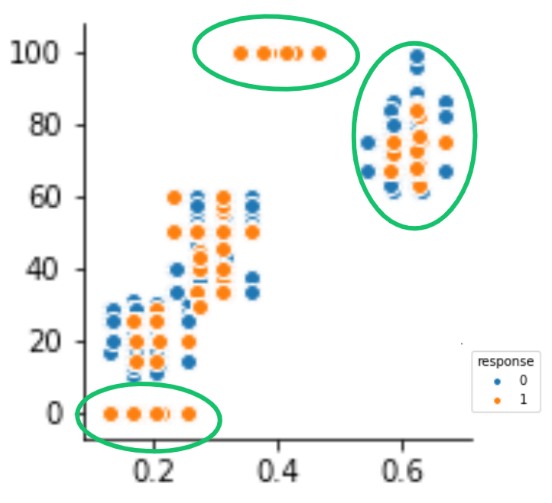
Now that you have some data in hand it is a good best practice to check that your data is in line with the underlying assumptions of your chosen machine learning model.

With many statistical tools, you can calculate things like the mean, inner-quartile range (IQR), and standard deviation. These tools can give you insight into the scope, scale, and shape of the dataset.

**Data visualization :-**

You can use data visualization to see outliers and trends in your data and to help stakeholders understand your data.

Look at the following two graphs. In the first graph, some data seems to have clustered into different groups. In the second graph, some data points might be outliers.



PLOT – 1 PLOT - 2

PLOT – 1 :- Some of the data seems to cluster in groups

PLOT – 2 :- Some of the data points seem to be outliers

**Terminology**

* **Impute** is a common term referring to different statistical tools which can be used to calculate missing values from your dataset.
* **Outliers** are data points that are significantly different from others in the same sample.