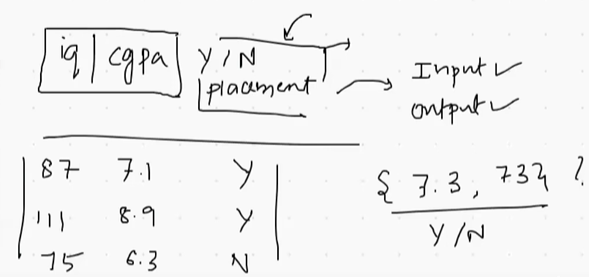
**DAY 2 [17/01/2022]: Types of machine learning**

**Machine learning types are:**

**1.Supervised Learning**

**If in the data, we have both input and output and our task is to find the relationship between input and output so that we can take new input and give the output, this learning is called Supervised Machine Learning. Most of the ML would belong to this category.**

**Example: Let’s take a data sample which consists of IQ, CGPA and placement status.**



**The above consist of both input (IQ, CGPA) and the output (placement status). Now ML will do the prediction and as in we take new input it would give us the output.**

**Sub categories in supervised learning:**

1. **Regression**

**Data are of two types.**

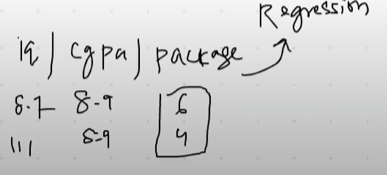
**1. Numerical- age, weight, marks**

**2. Categorical -gender, nation**

**In supervised machine learning, if the output data is of type numerical then it is Regression.**

**Example:**

**In the data given below, it consists both input and output, Also the output (package) is of numerical data.**



**Example: prediction of house price from various features.**

1. **Classification**

**In this type, output will be of categorical data as we have seen in the above supervised learning example.**

**Example: email spam classifier**

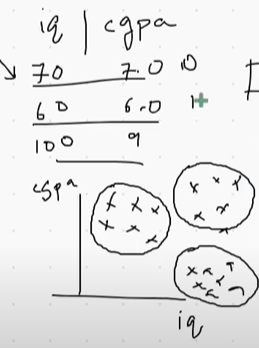
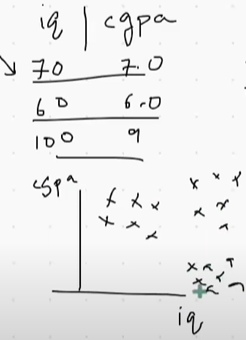
**2. Unsupervised Machine Learning**

**It consists only input not any output.**

**As the example mentioned for supervised ML, only IQ and CGPA would be given here in Unsupervised ML, from which we need to predict placement status.**

**Sub categories:**

1. **Clustering**



1. **(ii)**

**In the first figure, the given data of students has been plotted in a 2D coordinate system.**

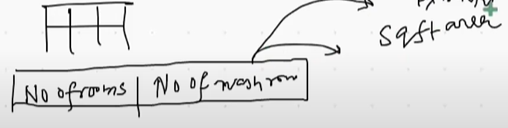
**Now the clustering algorithm will detect which students has to come in one group. We have categorized the students, as shown in second figure. Now we can label the data as 0- with high IQ and high CGPA, 1-with high IQ and low CGPA, 2-with low IQ and high CGPA, 3-with low IQ and low CGPA. And the main part is, the two columns on the data have been divided into three clusters. Hence n- dimensional data can be divided into clusters which cannot be seen through naked eye. Thus, it is a very powerful technique.**

1. **Dimensionality Reduction**

**When we deal with supervised machine learning, sometimes the data would be of image or textual based data which has about 1000’s input columns. Because of this large input columns,**

1. **the algorithm would run slow as we are working on large data**
2. **After a certain point, adding of new input columns won’t make a difference in result. There would be no necessity of some columns.**

**So, as the name says, dimensionality reduction will remove those extra columns.**



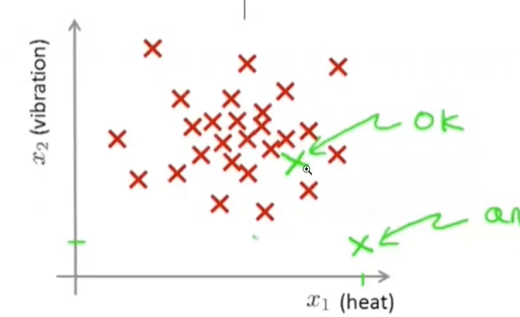
**As the above data of house price prediction contains two columns which is then converted into single column as square feet area. This is known as *Feature Extraction*. From the given multiple columns, we have created new column known as dimensionality reduction. We use**

**One more aspect is *Visualisation.* The data with multiple dimensions cannot be plot in a 3D coordinate system (graph) since in graph maximum of 3 dimensions can be shown. Hence what we do is, using dimensionality reduction algorithm we reduce the dimensions into 2D or 3D and in that coordinate ystem we plot those same points.**

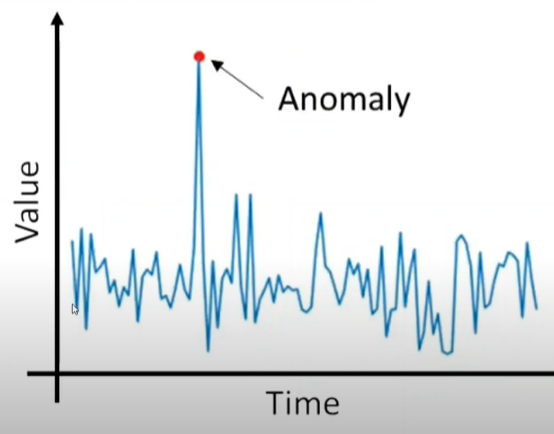


1. **Anomaly Detection**

**It used for detecting the outliers and remove it from the system such as credit card processing defect, manufacturing defect, loan approval problem, etc.**



**In the above graph, there are two green points. The green point which in middle of other red points can be considered as it won’t make a big difference. But the green point which is far from other points is said to be an anomaly. It may be classification as shown above or regression as shown below.**



1. **Association Rule Base Learning**

**Here we fetch the data and draw conclusions.**

**For example, in a supermarket, the problem is arranging the things in a order when we have 100’s of items. For this, let’s take out the bills of past two years. From this association rule, we will get to know milk and egg has been sold highest, hence we would keep it together. In this way we use this learning method.**

**3. Semi Supervised Learning**

**It’s been used for labelling data points. It’s a technique where we need not to label all the data points. Instead, we will label one or two data points and simultaneously other data points will be labelled by itself.**

**Example is Google Photos. As we click any photo of a person, it will identify his/her face in all other photos and combine it to a group. If we label one photo as friend then by default all other photos in that group would be labelled as friend.**

**4. Reinforcement Learning**

**As we know, supervised learning has both input and output, unsupervised learning deals has input but, in reinforcement learning, we will not have any data. The algorithm needs to start from the scratch and will improve later same as like humans. Ex: car driving**