← Go Back to Machine Learning

Course Content

Correlation vs Causation

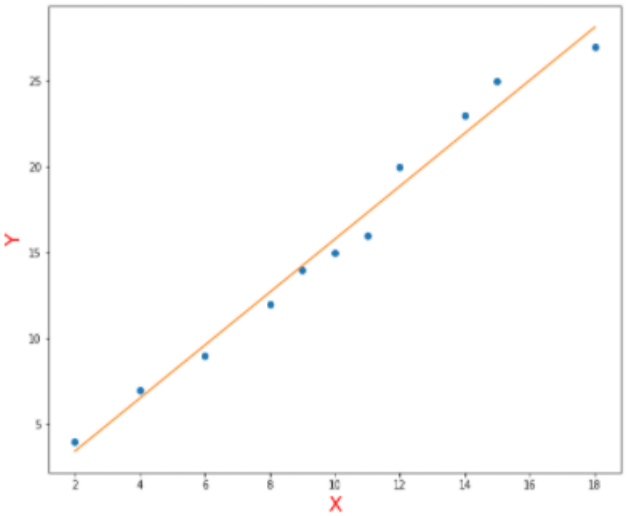
Correlation

The term correlation means that there is some kind of relationship or pattern between two variables.

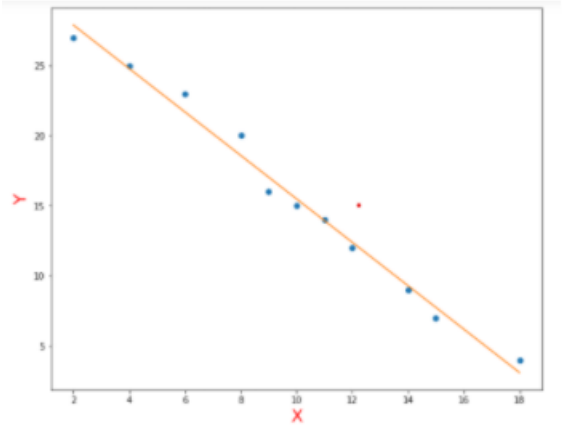
How can we determine if there is a correlation between variables?

Graphically, the correlation can be observed using a scatter plot between two variables.

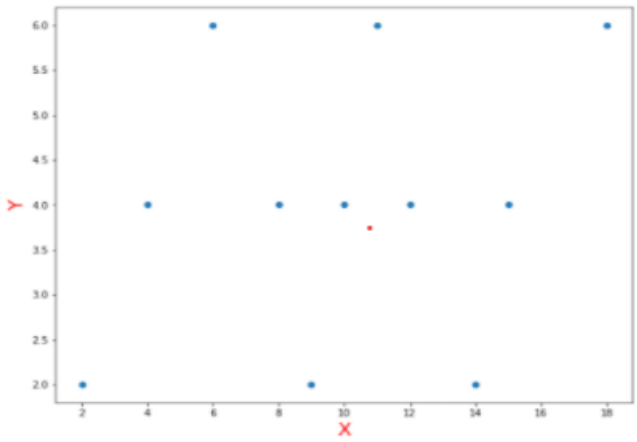
Positive correlation: When one variable increases, the other variable also increases in tandem. The below scatter plot shows a positive correlation between two variables.



Negative correlation: When one variable increases, the other variable decreases. The below scatter plot shows a negative correlation between two variables.



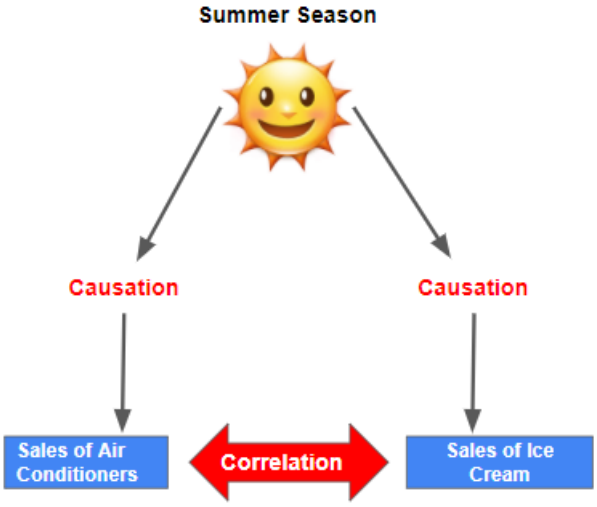
No (or zero) correlation: If the scatter plot shows that two variables are randomly scattered, i.e., one variable neither increases nor decreases with the other variable, then they may be thought to have no correlation between them.



Causation

Causation means that one event causes another event to occur.

Let's consider the example in the below image:



We know that in summers, there is likely to be an increase in sales of ice cream and sales of air conditioners. Hence, we can say that:

. There is a causal relationship between the summer season and sales of ice cream.
. There is a causal relationship between the summer season and sales of air conditioners.

Correlation vs Causation

It is important to note that correlation does not imply causation. There might be a third variable affecting the relationship between the two variables.

For example, we observed above that due to the causal relationship of summers with sales of air conditioners and sales of ice cream, the plot of sales of air conditioners and sales of ice cream in summers might show some positive correlation as both increase in the summer season but there is no particular causal relationship between the two variables.