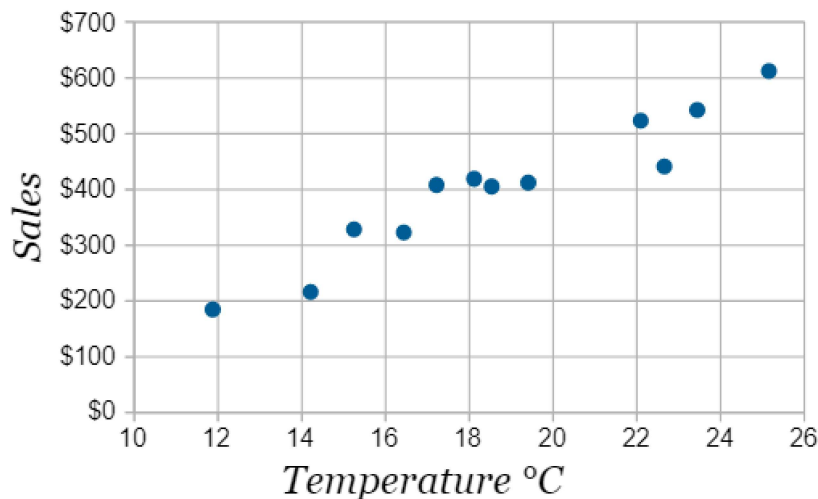


Module 2-1: Linear Regression

Collecting data is an important part of measuring trends and predicting results. There are many different types of data that can be collected such as weather reports, ice cream preferences, and clinical trials for medicine. Data can be gathered in many ways. The most popular methods are interviews, questionnaires, surveys, observations, documents and records.

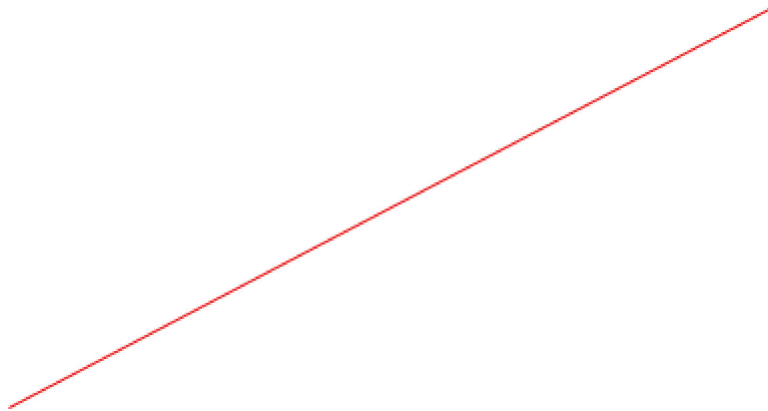
Weather reports can be obtained by observing weather instruments and previous year's weather records to see what the weather will be like today and to forecast for the week. Ice cream preferences can be gathered by surveying a population to help an ice cream shop market the most popular flavors. Data from clinical trials can be obtained by interviewing the candidates who took the new medicine to see how it affected them.

This is a basic graph of ice cream sales throughout the year. The x-axis is the temperature of the weather and the y-axis is the money made from sales. Each point on the graph represents a month of the year. For example, July is the hottest month of the year and so you can see the plot point of (25, 610) to be where the month of July plot point is plotted on this graph.

[Send Issue](#)

There is a very noticeable pattern in the plot points where the higher the temperature the higher the ice cream sales become. By looking

at the relationship between the temperature and the sales, a line can be drawn through the points to best represent all the points in the graph. We call this the **line of best fit**. The points don't form a line, but they kind of do.



From previous math classes learned in school the equation of a line is:

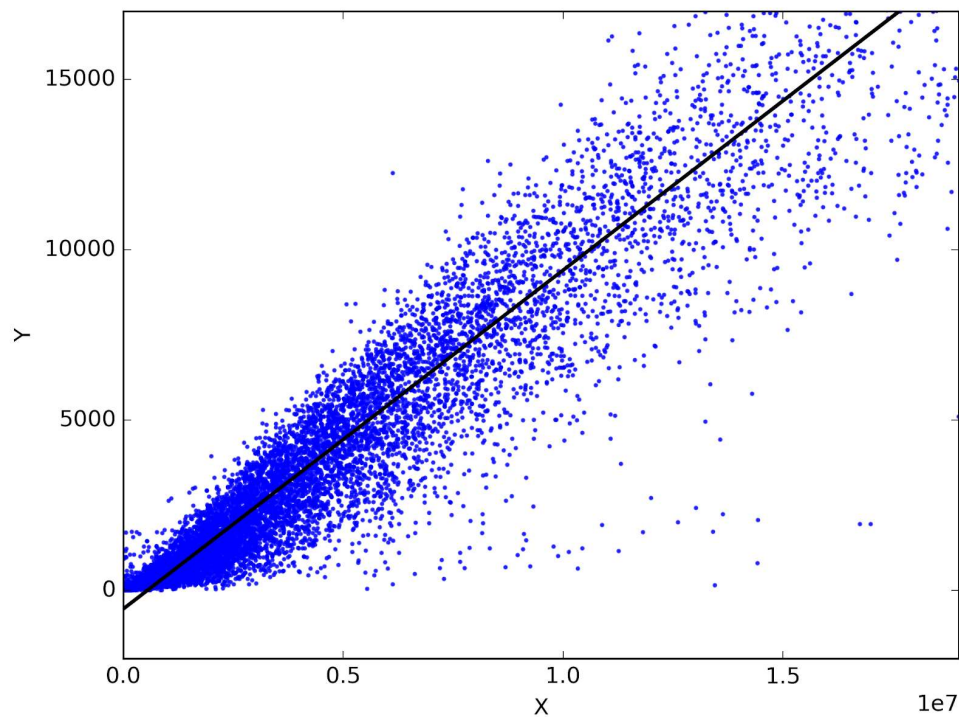
Text

COPY

$$y = mx + b$$

By finding out the linear relationship between the x and the y, we can predict what the point might be without having all the actual values. This process is known as **linear regression**. In later lessons, you will notice that the x-axis will be your features and the y-axis will be what you are predicting.

Here's another example of creating a line from data points. Notice again, how we can vaguely draw a line through the points and generally predict where points will be based on the line.



Creating a Linear Regression Program

Before starting on the linear regression program, to review, the steps for creating most artificial intelligence programs include:

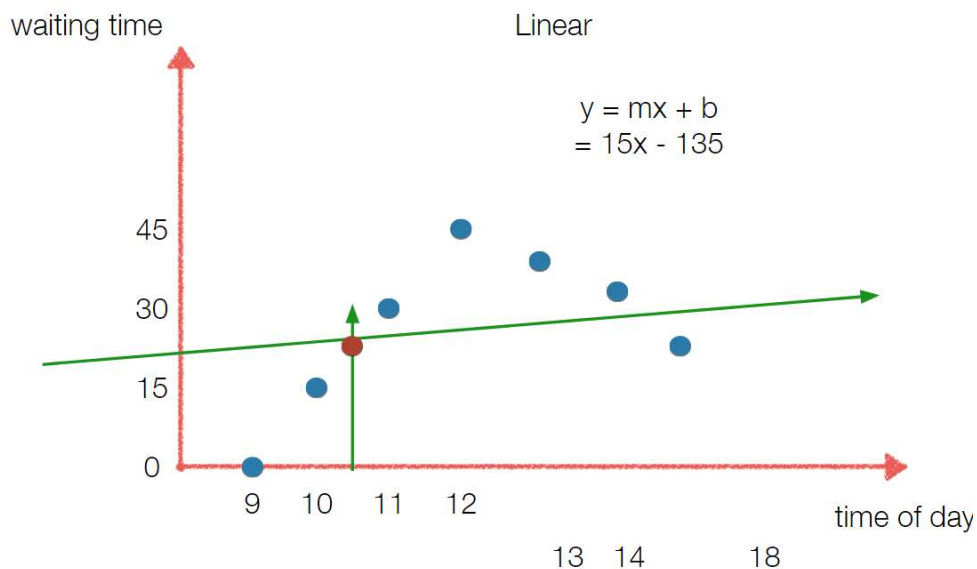
1. Define your data table
2. Collect your data
3. Choose a model
4. Machine Learning (fit)
5. Prediction

In this example of Disneyland Space Mountain Ride, we are trying to predict the waiting time of the ride based on many different features such as day of week, holiday, weather, and time of day.

The first two steps is to define your data table and collect your data. Here a "Day of Week" value of 1 means Monday.

When collecting your data there is input and output. The input is the features that ultimately affect the output. In our example, the blue will affect the red. The day of the week and time of day will change the wait time of the ride.

Day of Week	Holiday or Not	Weather	Time of the Day	Waiting Time
1			9	0
1			10	10
1			11	20
1			12	30
1			13	40
1			14	30
1			15	20
1			16	10



Now that you have defined your table and collected your data, you can start on the next step, which is to choose a model. We will be using the linear model along with the linear regression algorithm that was shown before.

To begin, you will need to import the `sklearn` library.

Python

COPY

```
from sklearn import linear_model
```

The next part is to create your input and output from the table that was created before. The input will be in the form of a list within a list and the output will be a list. For each sublist, the first value is the day, and the second value is the time of day. The order of the data matters, so make sure to double check with the table to have the input and output correct.

Text

COPY

```
input_data = [  
    [1, 9],  
    [1, 10],  
    [1, 11],  
    [1, 12],  
    [1, 13],  
    [1, 14],  
    [1, 15],  
    [1, 16],  
]  
output_data = [0,10,20,30,40,30,20,10]
```

The next step is machine learning, which is training our model on the input and output above.

Python

COPY

```
model = linear_model.LinearRegression()  
  
model.fit(input_data, output_data)
```

This creates the linear model that we will be using and specify that we will be using the Linear Regression algorithm. There are many different types of models and algorithms used in artificial intelligence, so specifying both is required.

The fit is used to fill the input and output data above into the model that we have created. The input data is always first and then the output data is after.

The last step is to predict an outcome using our model. The use of `predict` is needed for this.

Python

COPY

```
print(model.predict([[1, 11.5]]))
```

This will predict the wait time of the ride based on the data that was inside the list within a list and print it out to the display. To predict.