Supervised Regression Models - Linear Regression

By: eng.Esraa Madhi

What is Machine Learning?

So we can define machine learning as: Finds patterns in data and uses them to make predictions.
Machine learning model is: a mathematical representation of the patterns hidden in data.

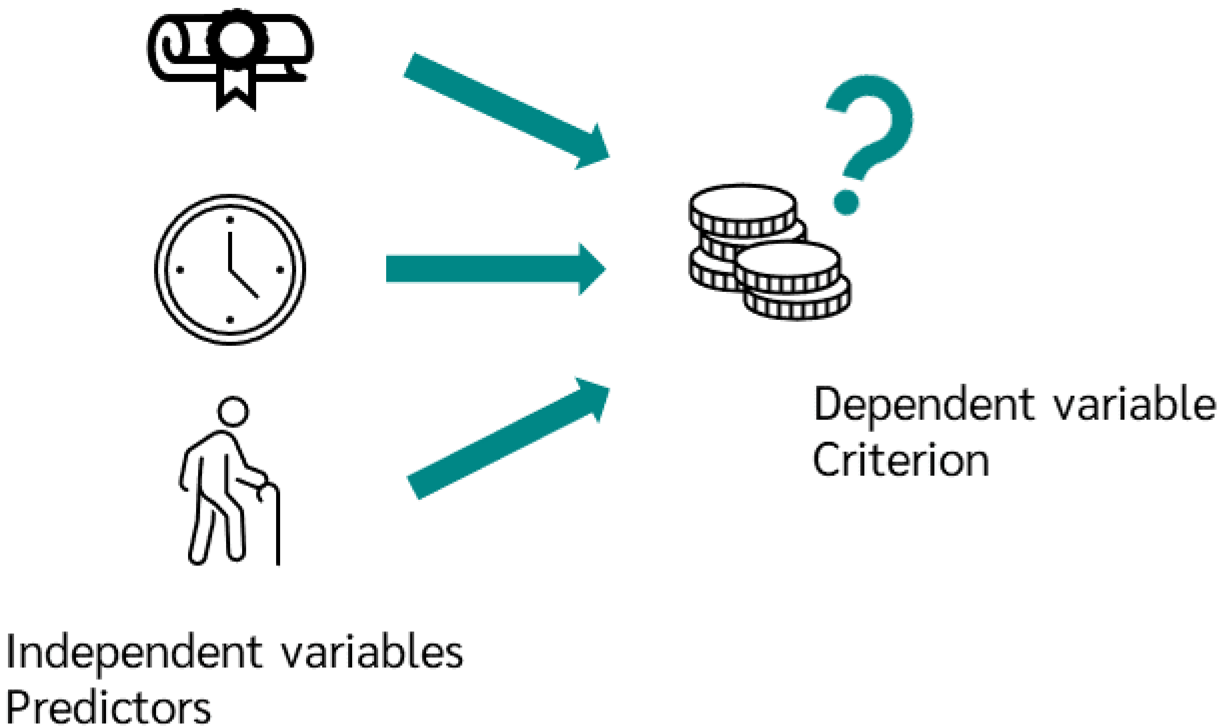
What is linear regression?

Linear regression identifies the relationship between the value of one variable (Continuous) and the corresponding values of one or more other variables

At its core, linear regression can help determine if one explanatory variable can provide value in predicting the outcome of the other. For example, does ad spending on one medium or another have any meaningful impact on sales?

Variables?

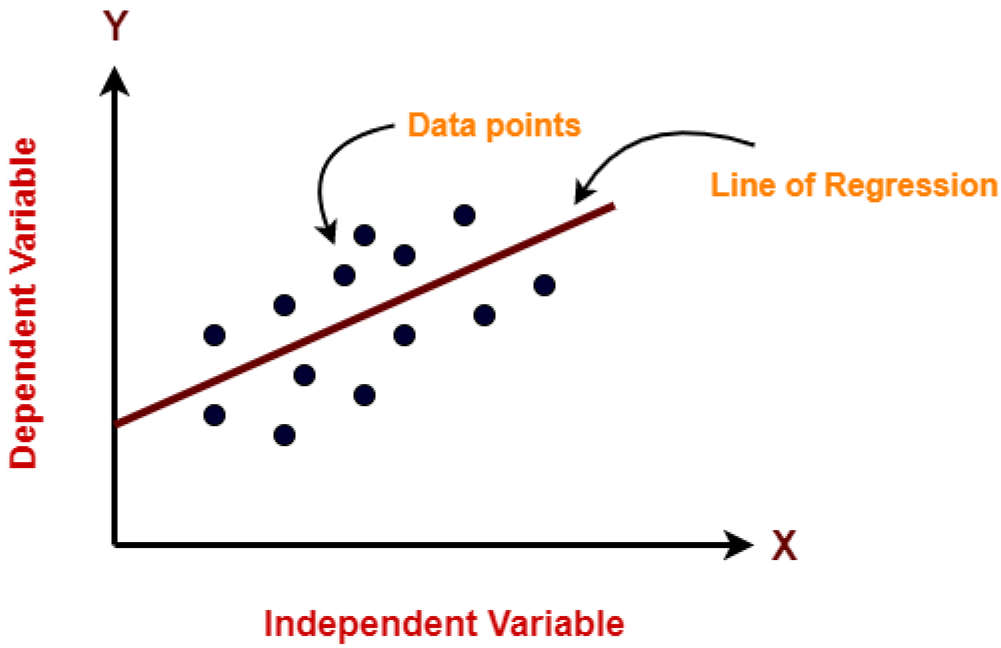
When linear regression tries to predict the value of one variable, called the *dependent variable*, given another variable, called the *independent variable*. For example, if you were trying to predict salary of employee based on his qualifications, salary would be the dependent variable, while (age, experience, number of certificates)would be the independent variable.



Let’s write the definition of linear regression again ,it is a statistical method used to predict the value of the dependent variable based on the values of the independent variables.

How linear regression finds this relations?

The fundamental concept to model the relationship between variables using various ML techniques to generate a regression line between variables such as sales rate and marketing spend.



This line also minimizes the difference between a predicted value for the dependent variable given the corresponding independent variable.

Linear regression assumes a linear relationship between the variables, and the **model is represented by a straight-line equation.**

What are linear regression types?

1. **Simple linear regression**

Simple linear regression finds a function that maps data points to a straight line onto a graph of two variables.

2. **Multiple linear regression (multivariate regression)**

Multiple linear regression finds a function that maps data points to a straight line between one dependent variable, like ice cream sales, and two or more independent variables, such as temperature and advertising spend.

3. **Polynomial regression**

A form of regression analysis in which the relationship between the independent variable and the dependent variable is modeled as an nth degree polynomial.

Linear Regression from mathematical perspective:

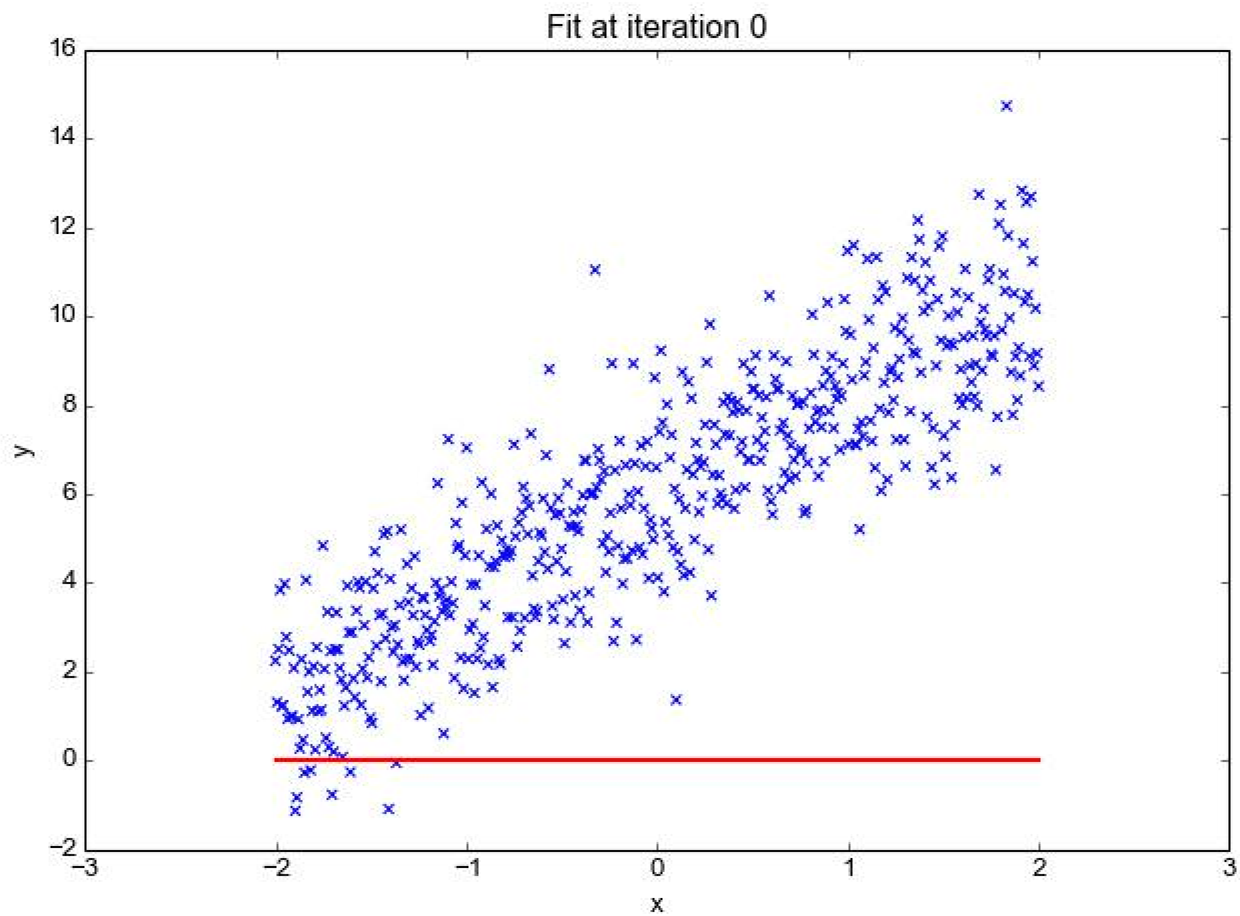
- The general form of the linear function for simple linear regression is:

```
y` = b0 + b1*X1
# where:
# y` is the dependent variable
# x1 is the independent variable
# b0 is the y-intercept (Where our line starts, indicating the basic y` when the X1 is 0)
# b1 is the slope (Shows how much our y` increase with a one-unit increase in X1
```

- For multiple linear regression, the function is similar but with multiple independent variables:

```
y` = b0 + b1X1 + b2X2 + … + bn*xn
# where:
# y` is the dependent variable
# x1, x2, …, xn are the independent variables
# b1, b2, …, bn are the respective regression coefficients.
# b0 is the y-intercept
```

The goal of linear regression is to find the values of the coefficients (b0, b1, b2, ..., bn) that minimize the difference between the predicted values of y and the actual values of y.



The model is → The Line	
The line is → the best coefficients makes the line fit the data	
The process to get best coefficients → Training	
The line fit the data → when the error is minimized as much as possible	
The error is minimized → when minimizes the difference between the predicted value for the dependent variable and the true value	

How to find best coefficients that get minimum error?

We need:

1. A way to calculate error
2. A way to reduce this error

A way to calculate error is → Loss Function
The loss is the error in our predicted value of target. Our goal is to minimize this error to obtain the most accurate value of the target
We will use the Mean Squared Error function as now to calculate the loss.

$$E = \frac{1}{n}\sum_{i=0}^{n}(y_i - \overline{y}_i)^2$$

$$E = -\frac{1}{n}\sum_{i=0}^{n}(y_i - (b1x_i + b0))^2$$

A way to reduce this error → we need an optimization method used to minimize the loss function. it's important to know there are different methods we could use it here.
One of them is:

Gradient descent is an iterative optimization algorithm to find the minimum of a function

How **Gradient descent** helps here? it estimates how we should change the coefficients value to get best fitting for the line in the data

The algorithm works as follows:

1. Initialize the values of b0 and b1 to random values.

2. Calculate the predicted values for the given input data using the current values of b0 and b1.

$$y_i = b1x_i + b0$$

3. Calculate the cost function using the predicted values and the actual values.

$$E = \frac{1}{n}\sum_{i=0}^{n}(y_i - \overline{y}_i)^2$$

4. Calculate the gradient (partial derivative) of the loss function with respect to b0 and b1.

$$D_{b1} = \frac{1}{n}\sum_{i=0}^{n} 2(y_i - (b1x_i + b0))(-x_i)$$

$$D_{b1} = -\frac{2}{n}\sum_{i=0}^{n} x_i(y_i - \hat{y}_i)$$

$D_{b0} = -\frac{2}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)$

5. Update the values of b0 and b1 by taking a step in the opposite direction of the gradient, with a step size determined by the learning rate.

6. Repeat steps 2–5 until the loss function is minimized or a maximum number of iterations is reached.

To illustrate how gradient descent works with a linear regression algorithm, let's consider a simple example dataset with one feature and a dependent variable. We'll go through one iteration of the algorithm using a single data point from this dataset.

Example Dataset

Suppose we have a dataset where the independent variable (x) represents hours studied, and the dependent variable (y) represents the score received on a test.

Hours Studied (x)	Test Score (y)
1	50
2	51
3	52
4	55
5	56

1. **Initialize the values of (b0) and (b1) to random values.**
   - Let's say (b0 = 0) and (b1 = 10) to start.

2. **Calculate the predicted values for the given input data using the current values of (b0) and (b1).**
   - For the first data point where $(x = 1)$, the prediction $(\hat{y})$ would be:

   $\hat{y} = b1 \cdot x + b0 = 10 \cdot 1 + 0 = 10$

3. **Calculate the cost function using the predicted values and the actual values.**
   - Using the squared error loss for this single data point where the actual $(y = 50)$:

   $\text{Cost} = (y - \hat{y})^2 = (50 - 10)^2 = 1600$

4. **Calculate the gradient (partial derivative) of the loss function with respect to (b0) and (b1).**
   - Gradient with respect to $(b1)$ is $((D_{b1}))$:

   $D_{b1} = -2 \cdot x \cdot (y - \hat{y}) = -2 \cdot 1 \cdot (50 - 10) = -80$
   - Gradient with respect to $(b0)$ is $((D_{b0}))$:

   $D_{b0} = -2 \cdot (y - \hat{y}) = -2 \cdot (50 - 10) = -80$

5. **Update the values of (b0) and (b1) by taking a step in the opposite direction of the gradient, with a step size determined by the learning rate.**
   - Assume the learning rate $(L = 0.01)$:

   $b1 = b1 - L \cdot D_{b1} = 10 - 0.01 \cdot (-80) = 10 + 0.8 = 10.8$
   $b0 = b0 - L \cdot D_{b0} = 0 - 0.01 \cdot (-80) = 0 + 0.8 = 0.8$

1. **Repeat steps 2–5 until the cost function is minimized or a maximum number of iterations is reached.**
   - For simplicity, we've only performed one iteration here. Normally, you would continue iterating, recalculating $(\hat{y})$, the cost, the gradients, and updating (b1) and (b0) until the changes in the cost function are minimal or after a set number of iterations.

This example shows the basic mechanics of applying gradient descent in a linear regression context with a very simple dataset.

https://youtu.be/jIfxXzjpmSc

https://youtu.be/jIfxXzjpmSc

What Advantages & Disadvantages of linear regression?

Advantages:
- It aids exploratory data analysis.
- It can identify relationships between variables.

- It is relatively straightforward to implement.

Disadvantages:
- It does not work well if the data is not truly independent.
- Machine learning linear regression is prone to underfitting that does not account for rare events.
- Outliers can skew the accuracy of linear regression models.

Linear Regression Assumption:

1. Linearity: Linear regression assumes a linear relationship between the dependent variable and the independent variable(s). This means that the relationship can be represented by a straight line.
2. Independence: Linear regression assumes that the observations are independent of each other. This means that there should not be any correlation among the observations.
3. No multicollinearity: Linear regression assumes that the independent variables are not highly correlated with each other.
4. Normality: Linear regression assumes that the errors (residuals) are normally distributed.

Violation of these assumptions can lead to inaccurate or unreliable results. It is important to check for these assumptions and address any problems before interpreting the results of a linear regression analysis.

Resources :
- https://utsavdesai26.medium.com/linear-regression-made-simple-a-step-by-step-tutorial-fb8e737ea2d9
- https://realpython.com/linear-regression-in-python/
- https://towardsdatascience.com/linear-regression-using-gradient-descent-97a6c8700931