Sanchit Minocha   Follow

Mar 31 · 7 min read

Linear Regression. Back to Basics.

You have been waiting for this since Day 1, ain't it? Because we like it unconventional, we saved it for the last. Next 2 days also cover *Regression*.

Linear Regression is the oldest, simple and widely used supervised machine learning algorithm for predictive analysis.

If she loves you more each and every day,
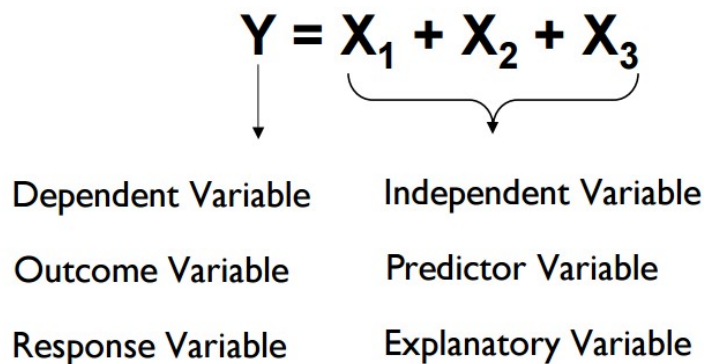by linear regression she hated you before you met.

Logic level - Infi.

Every Data Scientist starts with this one. So, here it is.

> *The path to your dreams is rarely linear. You gotta zigzag your way to happiness.*

.   .   .

What is Linear Regression?

It's a method to predict a **target variable** by fitting the *best linear relationship* between the dependent and independent variable.

Example!

What is the Best Fit?

It can be of any shape depending on the number of independent variables (a point on the axis, a line in two dimensions, a plane in three dimensions, or a hyperplane in higher dimensions).
Least Squares Method: The *best fit* is done by making sure that the sum of all the distances between the shape and the actual observations at each point is as small as possible. The fit of the shape is "best" in the sense that no other position would produce less error given the choice of shape.

If you want to go into the mathematics, you can have a look below or if you hate mathematics, just skip!

$$f(a, b) = a + b\,x,$$

so

$$R^2(a, b) \equiv \sum_{i=1}^{n} [y_i - (a + b\,x_i)]^2$$

$$\frac{\partial(R^2)}{\partial a} = -2 \sum_{i=1}^{n} [y_i - (a + b\,x_i)] = 0$$

$$\frac{\partial(R^2)}{\partial b} = -2 \sum_{i=1}^{n} [y_i - (a + b\,x_i)]\,x_i = 0.$$

These lead to the equations

$$n\,a + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i\,y_i.$$

$$b = \frac{\mathrm{cov}(x, y)}{\sigma_x^2}$$

$$a = \bar{y} - b\,\bar{x}.$$

Least Squares Method for Simple Linear Regression.

Note: I have shown this method only for Simple Linear Regression. You may extend it for Multiple Linear Regression. Don't know what Simple and Multiple Linear Regressions are? No problem, keep reading.

. . .

Application in Real Life

Pick any two things that you use in your daily life and that are related. Like, I have data of my monthly spending, monthly income and the number of trips per month for the last three years. Now I need to

answer the following questions:

What will be my monthly spending for *next year*?

Which *factor* (monthly income or number of trips per month) is more important in deciding my monthly spending?

How monthly income and trips per month are *correlated* with monthly spending?

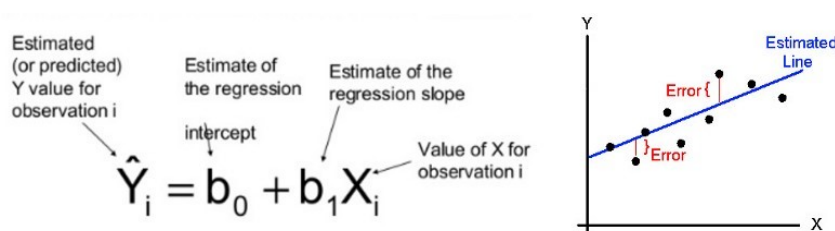**Yes, you are right. Linear Regression will come to your rescue.**

. . .

Types of Linear Regression

Yes, we are complicating it for you. Be open to **new** stuff. :D

1. Simple Linear Regression

This method uses a single independent variable to predict a dependent variable by fitting a best linear relationship.

2. Multiple Linear Regression

This method uses more than one independent variable to predict a dependent variable by fitting a best linear relationship.

In case of Multiple Regression, the parameters can be found in the same way as that in the case of simple linear regression, by minimising the cost function using:

• **Gradient Descent:** Given a function defined by a set of

parameters, Gradient Descent starts with an initial set of parameter values and iteratively moves towards a set of values that minimise the function. This iterative minimisation is done using calculus, taking steps in the **negative direction of the function gradient**.

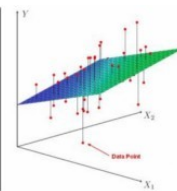$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Cost Function.

Repeat {
Gradient descent $\rightarrow$ $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \ldots, \theta_n)$
}   (simultaneously update for every $j = 0, \ldots, n$)

Gradient Descent.

Note: It works best when **multicollinearity** is absent. It's a phenomenon in which two or more predictor variables are highly correlated.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Multiple Regression for you.

3. Linear Splines

Sometimes, Linear splines is used to reduce the problem to Linear Regression. In this method, we fit the data with a *piece-wise linear function*. Let us suppose that the knots are at at k1 and k2 in the scatter plot as shown in the figures below. You may be thinking that we can divide the data into three groups using k1 and k2 and solve three regression problems (blue lines in left figure). But as you can see, **it does not assure continuity**!
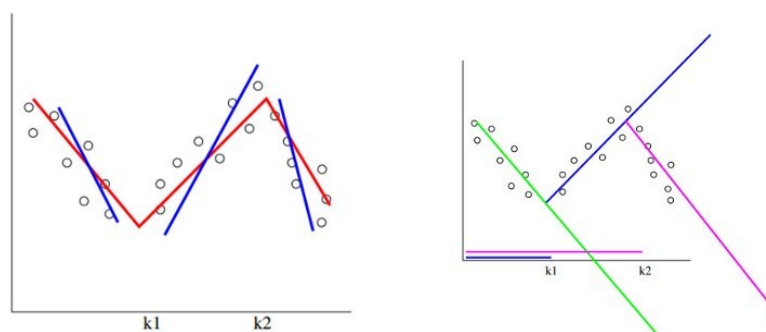
Example for Linear Splines.

To make the curve continuous , we can use the fact that any linear spline can be a linear combination of basis functions. Thus, the objective of the linear splines is to fit the red lines in the data (as shown in left figure). So, we can build a piece-wise linear function step by step.

We first start with a linear function for the points before k1 (green line in right figure).

$$\hat{f}(x) = a + bx$$

Next, we add up second function for the points between k1 and k2 (blue line in right figure).

$$\hat{f}(x) = a + bx + c(x - k_1)_+ \qquad (z)_+ = \left\{ \begin{array}{ll} z & \text{if } z \geq 0 \\ 0 & \text{if } z < 0 \end{array} \right.$$

At last, we will add a third function (purple line in right figure) for the points after k2.

$$\hat{f}(x) = a + bx + c(x - k_1)_+ + d(x - k_2)_+$$

4. Stepwise Regression

This regression model is used when we have **_more than one_** independent variable. It uses automatic procedure to select important independent variables and there is no human intervention.

Forward Stepwise Regression

Here, we start with **_null model_** which means it has no predictors, just one intercept (the mean over dependent variable).

Now, fit p (total number of variables) simple linear regression models, each with one of the variables in. Thus, we have just searched through all the single variable models, the best one and fixed this one in the model.

Similarly, search through the remaining **_p-1_** variables one by one, but this time with that variable in the model which was selected in previous step. Now choose the model which will be best among the p-1 models.

Continue until some stopping rule is satisfied like some **threshold value** of the number of variables to be selected.

Backward Stepwise Regression

It starts with the full least squares model containing all **_p_** predictors.

Now remove the variable with the largest p-value i.e. the least significant predictor.

The new model shall have (p-1) variables. Remove the variable with largest p-value again.

Continue until some stopping rule is satisfied like all variables have a p-value smaller than a **threshold value**.

So you can see, Stepwise Linear Regression is applying Multiple Linear Regression multiple times and selecting the important variables or removing the least significant predictors each time.

Note 1: For Backward Stepwise Linear Regression or Multiple Linear Regression to work fine, **the number of observations (n) should be**

more than the number of variables(p). It is because we can do least squares regression only when n is greater than p. For p greater than n, least squares model is not even defined.
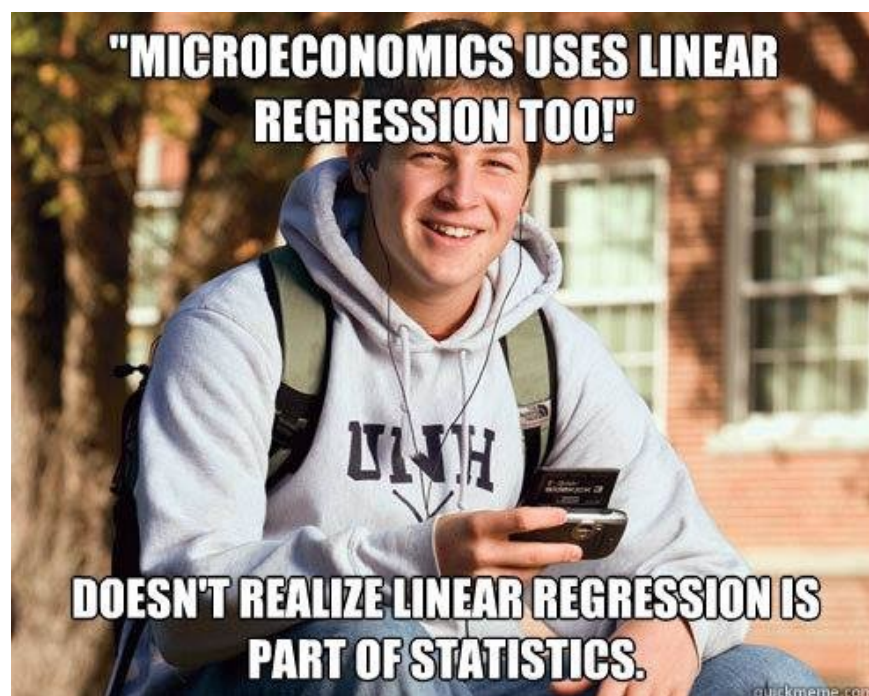
Note 2: Automatic procedures may not choose the right significant variables from practical point of view as they don't have the special knowledge the analyst might have.
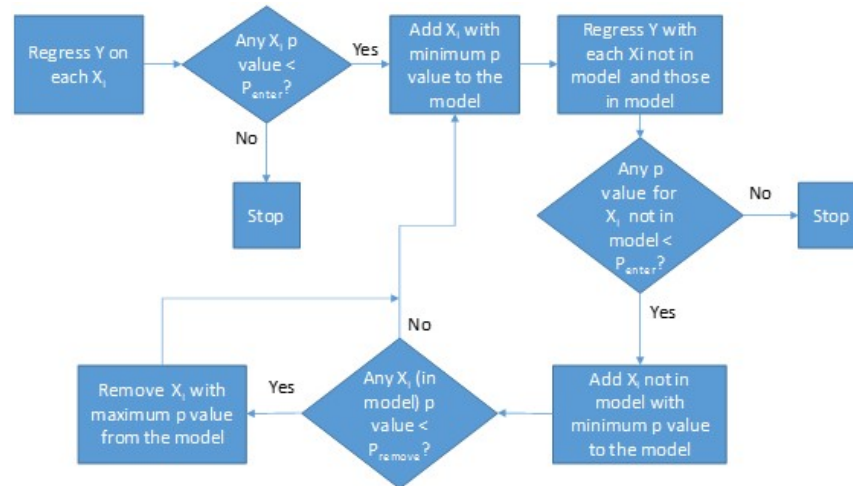
Stepwise Regression Model.

. . .

1. # How to select the right regression model?

Stats is very important. I repeat, VERY.

There is no perfect answer to this question because choosing the right linear regression model with a small sampled data is a difficult task. Well here, I will tell you some *common statistical methods* to get an idea which model best fits your data:

Adjusted R-squared and Predicted R-squared Value :- Choose the model with the **higher** adjusted and predicted R-squared values. Unlike, the adjusted and predicted R-squared values, which may increase or decrease on adding a predictor depending on the performance of model, regular R-squared value increases every time we add a predictor and may lead to overly complex model.

P-values for the predictors :- Variables with **low** p-values are the most significant variables.

Mallows' Cp :- It compares the precision and bias of the full model to models with a subset of predictors. The **smaller** it is, the more precise are the estimates of the true regression coefficients of the model.

Note: Apart from the above statistical methods, I would suggest that cross-validation is the best way to evaluate models.

Implementation in Python

So, it is important to see how to use what we have gained. I have applied Linear Regression on a self-created dataset in which I had recorded the *monthly income, monthly trips* and *monthly spending* for the last 3 years.

Click *HERE* for the full code and see how easy is the world of Machine Learning when it comes to Linear Regression.

References

1. Blog by Atomic Object

2. Minitab Blog

3. Stepwise Regression by Gerard Nico

Footnotes

Yes, this was a long one. But, the most important one too! We are almost standing at the end of 12A12D series. We hope you are learning.

Thanks for reading. :)
And, ❤ *if this was a good read. Enjoy!*

Editor: Akhil Gupta