

Developing Algorithms to Solve Real-World Problems

Mathematical models are used to solve a whole host of problems, from being able to better predict changes in the economy to the design of new drugs to fight diseases. In each case, complex numerical models are designed to fit real-world observations.

These models affect us all across our daily lives, so they must be designed accurately and efficiently.

Tyrone Rees, Nick Gould and Jennifer Scott, all part of Scientific Computing's Computational Mathematics group, have developed an algorithm that is capable of solving nonlinear least-squares problems. Least squares is a statistical method that is used to find the regression line (aka 'line of best fit') for a data-set collected from your real-world observations. This is a critical step in model design.

Tyrone tells us, "One of the central problems in computational mathematics is being able to fit a suitable model to observed data." The algorithm has now been incorporated into the Numerical Algorithms Group's library, an Oxford-based software company, where it has been used to develop a novel data fitting solver that is faster (solves 60% of problems in less time), more robust (solves 25% more problems) and requires fewer callbacks than its predecessor.

Significance of Least Squares in the Real World

Tyrone comments, "Models are used to represent, understand and simulate data that has been collected in an objective manner. When designed well, models allow us to form predictions and make logical decisions about the best course of action, according to the data."



Model design begins with a problem. The first step to designing a successful model is to understand what the fundamental nature of the problem is. Once this has been determined, we can use our current understanding of the situation to theorize a model that fits these defined first principles.

The next step is to establish the parameters – or the factors to be considered – for the model.

For example, in civil engineering, models are used to design roads in cities that reduce traffic.

The parameters for this may include:

- ◇ How many people drive?
- ◇ Where are the busiest driving spots?
- ◇ How many people commute from outside the city?
- ◇ Where do pedestrians walk?
- ◇ Where do pedestrians need to cross the road?
- ◇ Where are local business, shops and schools?
- ◇ Easy access for emergency services?

All of these factors must be taken into consideration to decide on the most efficient and safest road layout.



Once the parameters have been decided, they can be translated into mathematical symbols which can be solved, before being implemented into an algorithm. This is achieved by collecting the appropriate data. Once collected, it is important to establish if a relationship exists between the variables, by finding the regression line with minimal error. Least squares helps us do this.

If a relationship is found, it can be either linear or nonlinear. The difference between the two types of models is that the parameters of a linear model can be estimated in a linear manner, whilst the nonlinear models do not. The independent variables of linear models do not have to follow this rule to still be linear. This means that linear regression lines do not necessarily have to be straight lines. They can express different curves if the parameters are linear in nature, but an independent variable is raised to an exponent function, or it contains a logged or inversed term.

In real-world scenarios, most relationships are nonlinear in nature. For example, with regard to road design, how many people drive is not a constant variable. We see that there are peak traffic times each day, with more cars on the road at commuter times when driving to and from work or school. Even these daily peaks and troughs are not constant as we also have increases during holiday seasons, such as when people are returning home for Christmas.

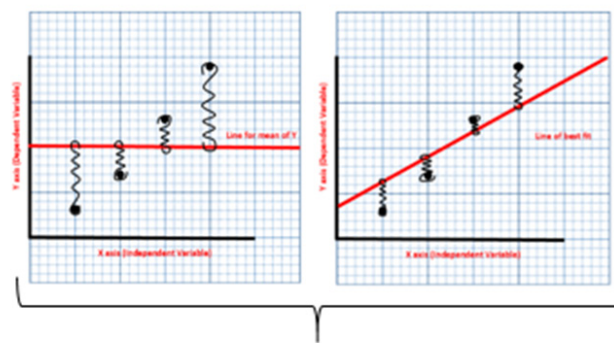
The Least Squares Method

To visualise this method of regression fitting, imagine you have collected your data and plotted it on a graph.

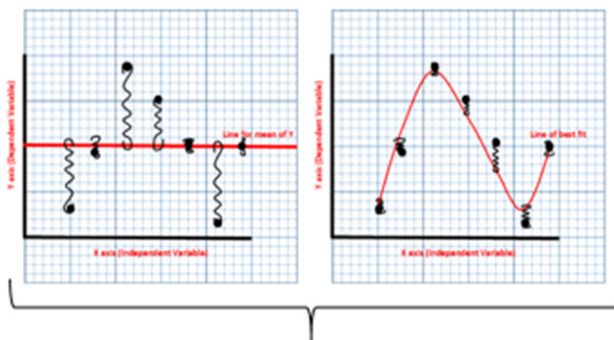
On this graph, you draw a line through the mean of Y (the average result for your dependent variable) and fix it in place. Then, you attach springs to the individual data points and the line you have drawn. You notice that some springs are more extended than others.

Now, if you allowed that line to move, the springs would exert a force on the line, rotating or bending it, and shortening the length of the springs.

By shortening the springs, we have found the regression line that has the least amount of error between our real-world observations and the model we have designed.



Linear Regression fitting



Nonlinear Regression fitting

In essence, the shorter the springs, the less error there is, and the better the model.

So, why use an algorithm?

Most models have a huge number of parameters to consider. Mathematicians consider a model fitting problem to be small to medium in size if it includes up to the thousands of parameters. In such cases, attempts at drawing a graph and regression line would be futile. This is made even more challenging when the relationship between the parameters is nonlinear in nature. So, to solve this issue, we use algorithms and software to determine the regression line for us. This allows us to design better models and make more informed decisions.

The algorithm designed by Scientific Computing's Computational Mathematics group was funded by an EPSRC grant.

Written by Hal Dodman
Designed by Dascha Choudhury-Gibbons
STFC Industrial Placement Students