Linear Regression | Regression (statistics) | Statistics (academic discipline) | Machine Learning

What do we optimize in linear regression?

3 Answers

John Fowler, works at California Institute of Technology
Answered Jun 9, 2017 · Author has **105** answers and **70.6k** answer views

In the most common forms of linear regression, “optimizing” is actually *minimizing*. When our data include uncertainties and the errors are sufficiently approximated as Gaussian random variables (the vast majority of cases in the physical sciences), we minimize the value of chi-square, which is defined as the sum over all data points of terms with the squared difference between the dependent data point and the fit in the numerator and the uncertainty variance for that data point in the denominator (this assumes that the errors are independent; the definition of chi-square is a little more complicated when they are not, and one needs an error covariance matrix).

If we have no prior uncertainties, then we minimize the sum of the squared differences between the data and the fit at each point.

One writes out the formal expression for this sum and sets the partial derivatives with respect to each fitting coefficient to zero. This yields as many equations as there are fitting coefficients, so one has a system of simultaneous equations that can be solved for the fitting coefficients, assuming no pathologies are involved.

So the “optimization” involves minimizing some form of summed squared deviations between the data and the fit. It can happen that other cost functions may be appropriate, such as the maximum deviation between the data and the fit, in which case the minimization takes a different form, but that is one of many details that are beyond the scope here.

This assumes that a mathematical model has been selected. I suppose that one could view selecting the best model as a form of optimization also, e.g., avoiding overfitting and underfitting by designing one with a sufficient but not excessive number of degrees of freedom, etc. Whereas there exist methods of automating the testing of different models, in my experience good judgment on the part of the analyst is the key ingredient.

Related Questions

How would you describe the difference between linear regression, lasso regression, and ridge regression?

What happens when we introduce more variables to linear regression?

Why is binning used in logistic regression but not in linear regression?

Is it possible to perform linear regression on unsupervised data?

How can I run linear regression in parallel with R?

How do we test endogeneity in linear regression model?

What does "linear function" in linear regression mean exactly? How is it different from polynomial regression?

Why does the linear regression work?

Is machine learning just glorified linear regressions?

Why is it called Linear Regression?

+ Ask New Question

Related Questions More Answers Below

How would you describe the difference between linear regression, lasso regression, and ridge regression?

What happens when we introduce more variables to linear regression?

Why is binning used in logistic regression but not in linear regression?

Is it possible to perform linear regression on unsupervised data?

How can I run linear regression in parallel with R?

+ Ask New Question

Frank Puk, PhD Operations Research & Machine Learning, The University of Texas at Arlington (2018)

Answered Sep 21, 2018 · Author has **165** answers and **152.8k** answer views

I assume you are talking about mathematical optimization for solving linear regression.

Indeed, linear regression can be solved in other ways including linear algebra, QR matrix decomposition, etc.

The formulation of solving a linear regression is as follows:

$$\min_{\boldsymbol{x}} \quad ||A\boldsymbol{x} - b||_2$$

As the formulation is convex, it can be solved by any standard QP prackages such as Cplex, Gurobi, CVX, etc.

However, the solution arrived by the above formulation is not necessarily unique. Therefore, we normally add one more regularization term as follows:

$$\min_{\boldsymbol{x}} \quad ||A\boldsymbol{x} - b||_2 + ||\boldsymbol{x}||_1$$

The above formulation is known as lasso (least absolute shrinkage and selection operator). Doing so can encourage sparsity on the solution vector and avoid over-fitting.

Reference: Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the lasso". Journal of the Royal Statistical Society. Series B (methodological). Wiley. 58 (1): 267–88.

461 Views

Helene Hoegsbro Thygesen, Ph.D. Biostatistics, University of Amsterdam (2006)

Answered Jun 9, 2017 · Author has **1.3k** answers and **884.9k** answer views

We minimize the sum of squared residuals. This is equivalent to maximizing the likelihood function, i.e. we find the fit that make our observations more likely than any other fit would.

822 Views · View 3 Upvoters

Related Questions

How do we test endogeneity in linear regression model?

What does "linear function" in linear regression mean exactly? How is it different from polynomial regression?

Why does the linear regression work?

Is machine learning just glorified linear regressions?

Why is it called Linear Regression?

What are some important parameters to evaluate linear regression results?

What is Logistic regression and what are its advantages over linear regression?

How can I prove the linear regression method?

What is the difference between a multiple linear regression and a multivariate regression?

What is fit_intercept in sklearn linear regression?

Why is the coefficient of the same regressor obtained from linear regression different from that obtained from multiple linear regression?

Can linear regression work on non-linear data sets?

Is linear regression a good trading strategy?

How can we overcome the limitations of linear regression analysis?

How do I report multivariate linear regression results?

+ Ask New Question