Comparison of Estimation Methods for Regression Coefficients

Olubayode Ebenezer

April 1, 2024

Abstract

This document presents a comparison of various estimation methods for regression coefficients, including closed-form Ordinary Least Squares (OLS), Gradient Descent, L-BFGS, and the Nelder-Mead algorithm. Through this analysis, we aim to highlight the effectiveness and appropriateness of each method in approximating the true regression coefficients under different circumstances.

1   Introduction

Regression analysis is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables. The accuracy of parameter estimation is crucial for the reliability of predictions made by regression models. This report compares the estimation accuracy of four different methods: closed-form OLS, Gradient Descent, L-BFGS, and Nelder-Mead algorithms.

2   Methodology and Results

2.1   True β Values

The true β values for our comparison are given as:

$$\beta = [1.5, -1, -0.25, 0.75, 3.5, -2, 0.5, 1, 1.25, 2]$$

2.2 Closed-form OLS Solution

The closed-form Ordinary Least Squares (OLS) estimates are very close to the true β values, indicating high accuracy. The estimates are as follows:

$$\hat{\beta}_{OLS} = [1.4995, -0.9995, -0.2497, 0.7495, 3.4996, -1.9992, 0.4996, 0.9989, 1.2510, 1.9988]$$

2.3 Gradient Descent

The Gradient Descent method yielded estimates significantly smaller in magnitude compared to the true b values. This indicates that the Gradient Descent method, with the given learning rate and number of iterations, did not converge to the true parameter values. What this implies is that such discrepancies highlight the sensitivity of the Gradient Descent algorithm to its hyperparameters and the importance of careful tuning, especially in cases where the parameter space is large, or the model is complex. For accurate estimation, one might need to experiment with different learning rates, increase the number of iterations, or consider alternative optimization strategies that are more resilient to initial parameter settings.

The estimates are:

$$\hat{\beta}_{OLS}GD = [0.0090, -0.0059, -0.0014, 0.0046, 0.0208, -0.0119, 0.0030, 0.0059, 0.0076, 0.0118]$$

2.4 L-BFGS Algorithm (OLS)

The L-BFGS algorithm estimates match closely with the true β values, similar to the closed-form solution. This showcases the efficiency of L-BFGS as an optimization method that combines the accuracy of analytical methods with numerical optimization's flexibility. The estimates are:

$$\hat{\beta}_{OLS}L-BFGS = [1.4995, -0.9995, -0.2497, 0.7495, 3.4996, -1.9992, 0.4996, 0.9989, 1.2510, 1.9988]$$

2.5 Nelder-Mead Algorithm

Estimates from the Nelder-Mead algorithm were less accurate, which explains its limitations in parameter estimation for regression, especially due to its heuristic, gradient-free approach which might not always converge optimally in high-dimensional spaces.

The estimates are:

$$\hat{\beta}_{OLS}NM = [0.7601, -1.2399, -1.0077, 0.4001, 2.6770, -2.6082, -0.2414, -0.7113, 1.0827, 1.5081]$$

2.6 L-BFGS vs. Nelder-Mead

The L-BFGS provided estimates nearly identical to the true β values, whereas Nelder-Mead's estimates were less accurate. I can say this difference shows the superiority of gradient-based optimization (L-BFGS) over heuristic, gradient-free methods (Nelder-Mead) for accurate parameter estimation.

2.7 MLE using L-BFGS Algorithm

The MLE estimates using the L-BFGS algorithm closely match the true β value

The estimates are:

$$\hat{\beta}_{MLE} = [1.4995, 0.9995, 0.2497, 0.7495, 3.4996, 1.9992, 0.4996, 0.9989, 1.2510, 1.9988]$$

3 Discussion

The precision of the closed-form solution and L-BFGS (for both OLS and MLE) is noteworthy, and the performance of the Gradient Descent approach emphasizes the importance of very precise hyperparameter adjustment.

4 Conclusion

This comparison highlights the critical aspects of each estimation method and their implications for statistical analysis and predictive modeling. It demonstrates the importance of careful method selection and hyperparameter tuning in achieving accurate parameter estimation.

5   Model Summary

Dep. Variable:	y	R-squared:	0.997
Model:	OLS	Adj. R-squared:	0.997
Method:	Least Squares	F-statistic:	4.375e+06
Date:	Sat, 30 Mar 2024	Prob (F-statistic):	0.00
Time:	23:23:44	Log-Likelihood:	-3080.3
No. Observations:	100000	AIC:	6181.
Df Residuals:	99990	BIC:	6276.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.4995	0.001	1900.039	0.000	1.498	1.501
x1	-0.9995	0.001	-1261.812	0.000	-1.001	-0.998
x2	-0.2497	0.001	-315.467	0.000	-0.251	-0.248
x3	0.7495	0.001	948.510	0.000	0.748	0.751
x4	3.4996	0.001	4425.582	0.000	3.498	3.501
x5	-1.9992	0.001	-2535.789	0.000	-2.001	-1.998
x6	0.4996	0.001	630.970	0.000	0.498	0.501
x7	0.9989	0.001	1263.691	0.000	0.997	1.000
x8	1.2510	0.001	1582.645	0.000	1.249	1.253
x9	1.9988	0.001	2529.516	0.000	1.997	2.000

Omnibus:	1.330	Durbin-Watson:	2.001
Prob(Omnibus):	0.514	Jarque-Bera (JB):	1.321
Skew:	-0.008	Prob(JB):	0.517
Kurtosis:	3.006	Cond. No.	1.02

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.