Huriye Güngör

78014

CSSM 502 Homework 2 Report

In this homework, I used iris dataset to practice linear regression model. The dataset consists of four features namely Sepal Length, Sepal Width, Petal Length, Petal Width and a target variable which is Class.

In this example, I would like to observe whether these features have positive or negative regression coefficients (β). Before I started to analyze it, I believed that they would have positive regression coefficients.

Feature	Regression Estimate (by hand)	Regression Estimate (by module)	Standard Error	Lower Credible Interval	Upper Credible Interval
Sepal Length (cm)	0.1829	0.7742	0.0099	5.7097	5.9769
Sepal Width (cm)	0.3047	-0.8019	0.0235	2.9870	3.1277
Petal Length (cm)	0.2974	0.4404	0.0075	3.4732	4.0428
Petal Width (cm)	0.8894	1.0281	0.0156	1.0764	1.3223

From the table above, it can be seen that the regression estimates that calculated by hand method and imported LinearRegression module has some discrepancies. The biggest difference arises from Sepal Width which is positive in by hand method, and negative in by module method. This may be due to the fact that Sepal Width has the largest standard error value among the others. Moreover, it can be seen that feature with the smallest standard error, Petal Length) has similar regression coefficients for two methods, compared to the other ones.

Finally, by looking at the 95% credible intervals, no upper limit and lower limit coincides with each other, which can also be used as a valuable information where we know a specific measure (cm); however, we do not know which feature it belongs to.