

SSY316 - Python 1

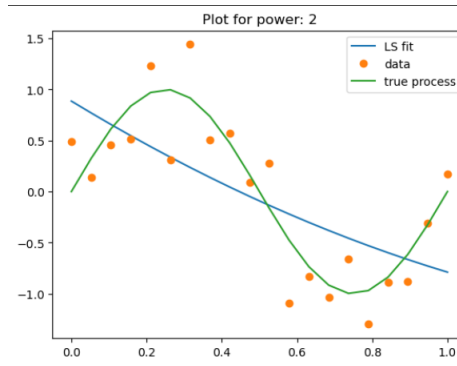
Fengyuan Liu Pär Aronsson

November 2023

1 Using Python's built-in methods

This assignment gave us insights into how one can estimate the parameters for linear regression by our own implementation and using Python's built-in functions.

Specifically for the first part, we learned how you could estimate the interception and coefficients for a regression model using MLE and OLS, which gave us the β values [0.88516511, -2.23262046, 0.55870444] which are seen in the figure 1. What was also interesting was how we were able to view the confidence intervals for our estimated β



2 Bayesian Linear Regression

The next part we learned from the assignment was estimating the σ^2 by starting off with a given prior from the Gamma distribution $\text{Gamma}(a_o, b_0)$ and getting the posterior $\text{Gamma}(a_n, b_n)$ by using the given formulas in figure 2. Also later sampling different σ^2 using Monte Carlo in order to estimate our β .

$$a_n = a_0 + \frac{n}{2}$$

$$b_n = b_0 + \frac{1}{2}(y^T y + \mu_0^T \Omega_0^{-1} \mu_0 + \mu_n^T \Omega_n^{-1} \mu_n).$$

3 Marginal likelihood / Model edvidence

The last section analyzes the marginal likelihood of our model parameters by writing the probabilities of our estimated β and σ^2 using these equations in figure 3.

$$\pi(\beta, \sigma^2 | y, X) = \frac{\pi(y | \beta, \sigma^2, X) \pi(\beta, \sigma^2)}{\pi(y | X)},$$

$$\pi(y | X) = \frac{\pi(y | \beta, \sigma^2, X) \pi(\beta, \sigma^2)}{\pi(\beta, \sigma^2 | y, X)},$$

A question asked was what order had the best fit and the result we got from it was that excluding order 1 it was order 2 which estimate was -28.696 and all the results can be seen in figure 3.

4 Finding the best fit for the Automobile dataset

The last task was to find the best parameters to fit the economic loss given three other variables. Using what we have learned we tried to find the best combination of the variables that would give us the best estimate of economic loss. Trying all combinations gave us the answer that 'LOSS' was the best fit which can be seen in figure 4. If we excluded loss as the best variable the answer was in that case 'ATTORNEY' and 'CLMAGE'

```
order: 1
-27.179883452794552
order: 2
-28.695588113412835
order: 3
-29.901259649596263
order: 4
-31.42057596554511
order: 5
-32.94083847245354
order: 6
-34.46291489713373
```

```
['ATTORNEY', 'CLMSEX', 'CLMAGE']
['ATTORNEY', 'CLMSEX']
['ATTORNEY', 'CLMAGE']
['CLMSEX', 'CLMAGE']
['ATTORNEY']
['CLMSEX']
['CLMAGE']
['LOSS']
-----
The best found model used these parameters:
-4477.815309796096
['LOSS']
```