

Regression analysis and resampling methods

Elisabeth Strøm

October 7, 2019

Contents

1	Introduction	3
2	Method	3
2.1	Implementation	3
3	Results	3
3.1	a)	3
3.2	b) Resampling	4
3.3	c) Bias-Variance tradeoff	4
3.3.1	1	4
3.3.2	2	4
3.4	d) Ridge regression	5
3.5	e) Lasso regression	9
4	Discussion	9
5	Conclusion	9

1 Introduction

2 Method

1σ confidence interval of β parameters for OLS, is, assuming a Gaussian distribution The interval has a confidence level of 95%

$$CI_X^{0.95} = E[X] \pm 1.96 \cdot STD_X, \quad (1)$$

since $E[\beta] = \beta$,

$$CI_\beta = \beta \pm 1.96 \sqrt{\text{Var}(\beta)}. \quad (2)$$

2.1 Implementation

3 Results

3.1 a)

Franke function, fit a polynomial to the fifth order.

Confidence interval of β -parameters, by computing the variances, evaluate the mean squared error(MSE), and R^2

The beta-parameters along with their confidence intervals are seen in figure 1

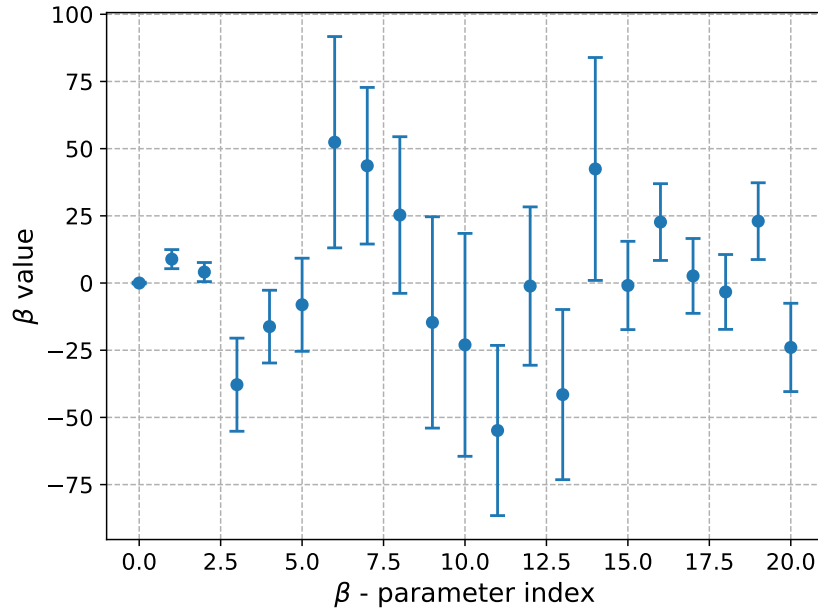


Figure 1: The values of the parameters β found from an Ordinary Least Squares fitting, are shown on the y -axis, while the index of the parameters are on the x -axis. The errorbars are calculated at 95% confidence

See a large spread in values, and some very large confidence intervals, some spanning from negative to positive values.

MSE score of $4.09 \cdot 10^{-3}$, and R^2 score of $9.50 \cdot 10^{-1}$

Want a low MSE, and a R^2 score close to 1.

Same result using sklearn

3.2 b) Resampling

Implement k -fold cross validation (CV). We use $k = 5$, and our test set is $\frac{1}{5}$ of the full dataset.

When calculating MSE using $\text{mean}(y_{\text{true}} - \tilde{y})$, $\text{MSE} = 4.62 \cdot 10^{-3}$, should do, when we know the true function as we do know. $R^2 = 0.944$

Assuming we did not know the true underlying function, and we use our noisy model instead, $\text{MSE} = 1.19$

which should be the irreducible noise, that is, σ^2 , which in our case we had chosen as 1.

3.3 c) Bias-Variance tradeoff

3.3.1 1

Rewrite cost function into bias+variance+irreducible noise

3.3.2 2

Figure similar to Fig. 2.11 of Hastie, Tibshirani, and Friedman , which shows the test and training error as a function of the complexity of the model. In our case, the complexity is the polynomial degree.

To reproduce this figure, we will pretend we do not know the true model, and use the noisy data alone. We then get the diagram in Figure 2. Lowest $\text{MSE} = 1.19$ and $R^2 = -0.095$

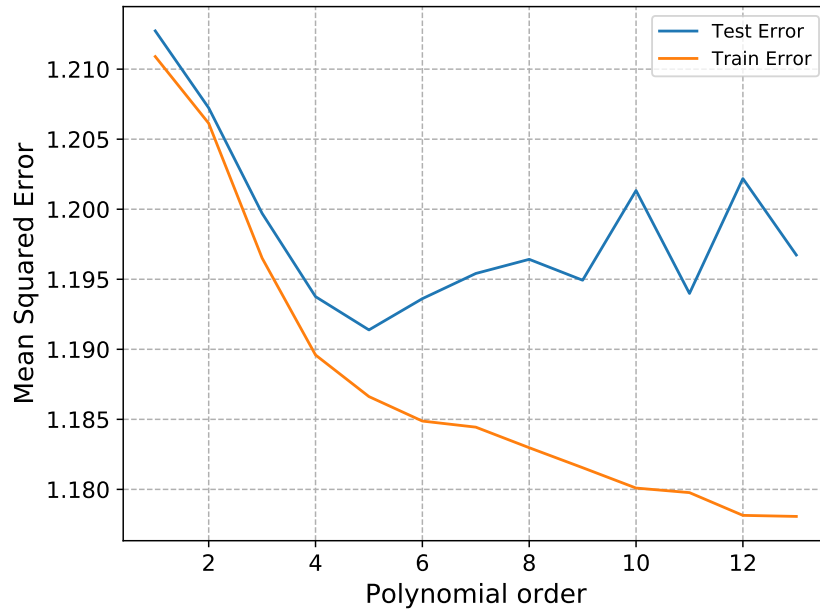


Figure 2: The training and test MSE for increasing model complexity, using Ordinary Least Squares and k -fold Cross Validation. While the test error reaches a minimum before increasing again, the training error will continuously decrease with increasing complexity.

Where the training error start going lower than 1, while the test error becomes larger than 1, is where the model start over-fitting, that is, fitting the noise. READ 2.9 OF HASTIE. INCLUDE BIAS AND VARIANCE PLOT IF TIME

Using the true model instead, Lowest $\text{MSE} = 4.51 \cdot 10^{-3}$ and $R^2 = 0.945$

3.4 d) Ridge regression

Perform the same analysis a a)-c).

Start then with fitting a fifth order polynomial to the data. First we must choose a λ value. We calculate the MSE for a fifth order polynomial fit, with several values of λ and using cross validation. This can be seen in Figure 3. We find that $\lambda = 10^{-3}$ minimizes the MSE.

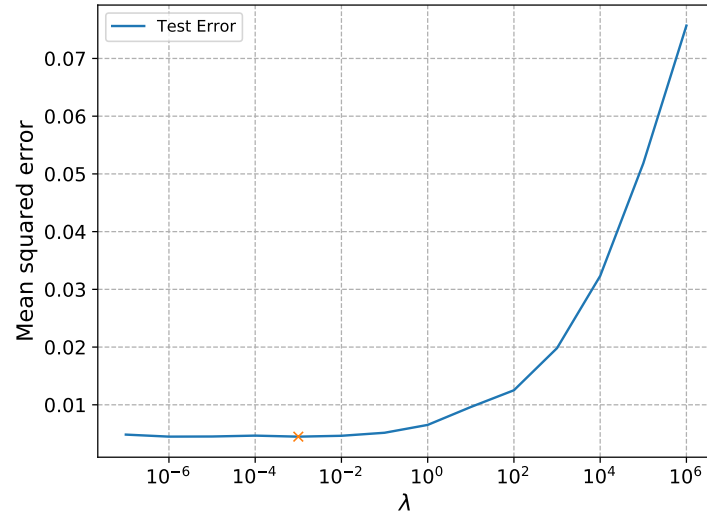


Figure 3: The MSE, calculated using the true model, for the test data for various λ values, while using cross validation. The minimum MSE value has been marked by a yellow \times .

Without cross-validating, using a fifth order polynomial and $\lambda = 10^{-3}$, the MSE score is $4.45 \cdot 10^{-3}$, and $R^2 = 0.95$

Using resampling, $MSE = 4.51 \cdot 10^{-3}$, $R^2 = 0.95$.

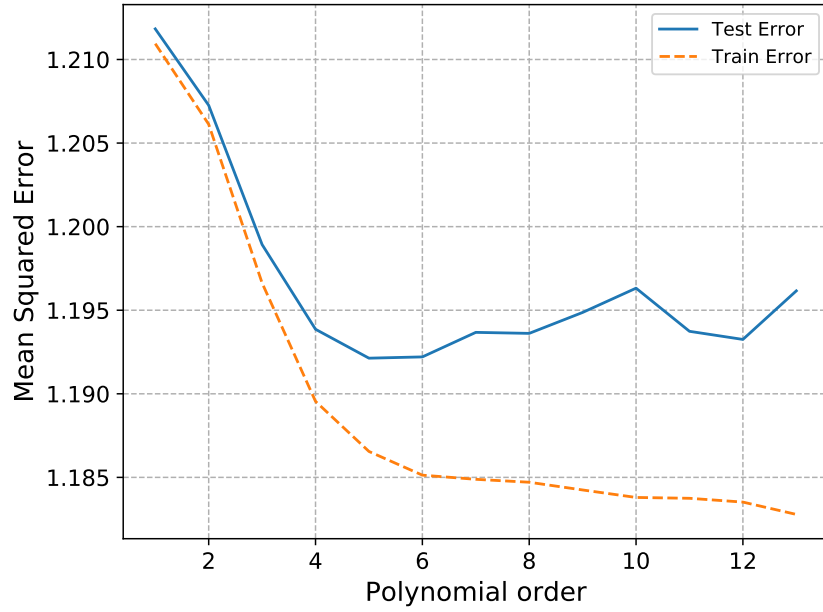


Figure 4: The training and test MSE for increasing model complexity, using Ridge regression and k -fold Cross Validation. While the test error reaches a minimum before increasing again, the training error will continuously decrease with increasing complexity.

Calculating MSE as both a function of λ and polynomial degree, we get the plot in Figure 5. According to the figure, a fifth order polynomial which $\lambda = 10^{-3}$ fits the data the best, but we see also that the increase in MSE is very small for increasing complexity and λ values between 10^{-5} to 10^{-1} .

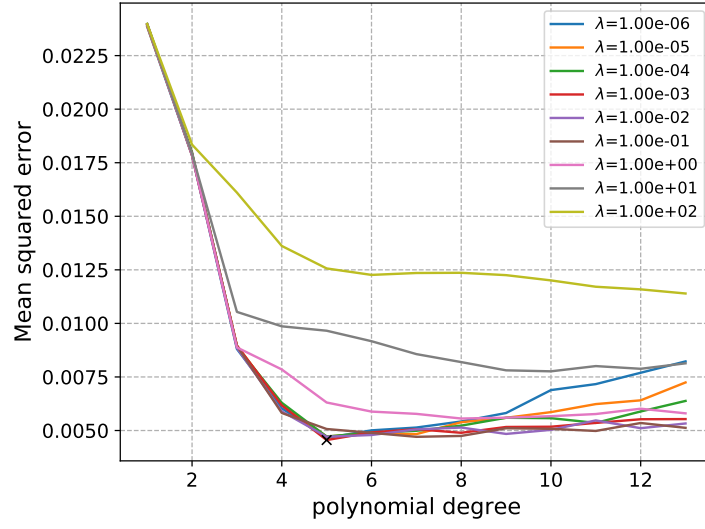


Figure 5: The MSE, calculated using the true model, for the test data for various λ values and polynomial degrees, while using cross validation. The minimum MSE value has been marked by a yellow \times .

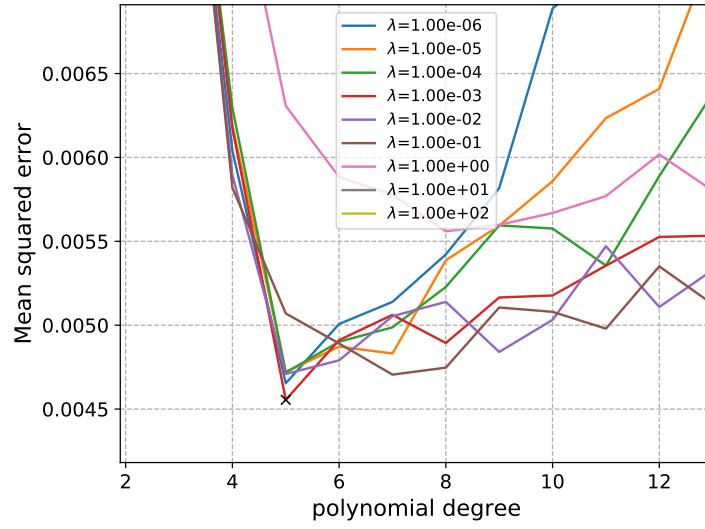


Figure 6: The MSE, calculated using the true model, for the test and training data for various λ values, while using cross validation, for a fifth order polynomial. The minimum MSE value has been marked by a yellow \times .

3.5 e) Lasso regression

Found that $\lambda = 10^{-7}$ fit data best, but so does any smaller value, does not seem to stop. $MSE = 4.45 \cdot 10^{-3}$

Found also that fifth degree polynomial fits the data the best, with the same λ

4 Discussion

5 Conclusion

References