Ridge regression

Daniel Eberharter
Claudio Canella

June 8, 2014

1 Algorithmic Choices

Our method does not substantially differ from the one discussed in class. We chose two non-linear functions, one is $f(x) = x^2 + 2x + 2$ and the other one is $g(x) = 6x^2 + 2x + 2$. We then determine our data sets by calculating $r^t(x) = f(x^t) + \epsilon$ (and $r^t(x) = g(x^t) + \epsilon$ respectively for $g(x)$), where ϵ is drawn from a standard uniform distribution with a random error selected between 0 and 15 weighted by the variance(for full information on calculation please see the octave script). The x-values are linear distributed between x_{min} and x_{max}
We calculate our design matrix and use the formula that has been introduced in the lecture and further determine the coefficients w_i for the regression function for each given value of λ (to avoid biasing towards w_0 the y-values are centered).
By using M-fold cross-validation we find the best λ out of the chosen values.

2 Numbers and Illustrations

- $x_{min} = 0, x_{max} = 9, x_n = 20$
- $M = 3$
- probability of error: 15% (used for ϵ)
- λ values are logarithmic equally distributed between $\lambda_{min} = 1$ and $\lambda_{max} = 10^6$, $\lambda_n = 19$
- The λ_i to be plotted are selected as follows (where b is the index of the best λ)

$$\lambda_b$$

$$\lambda_{1+(b-\frac{\lambda_n}{2})\%\lambda_n}$$

$$\lambda_{\lambda_n-(1+(b-\frac{\lambda_n}{2})\%\lambda_n}$$

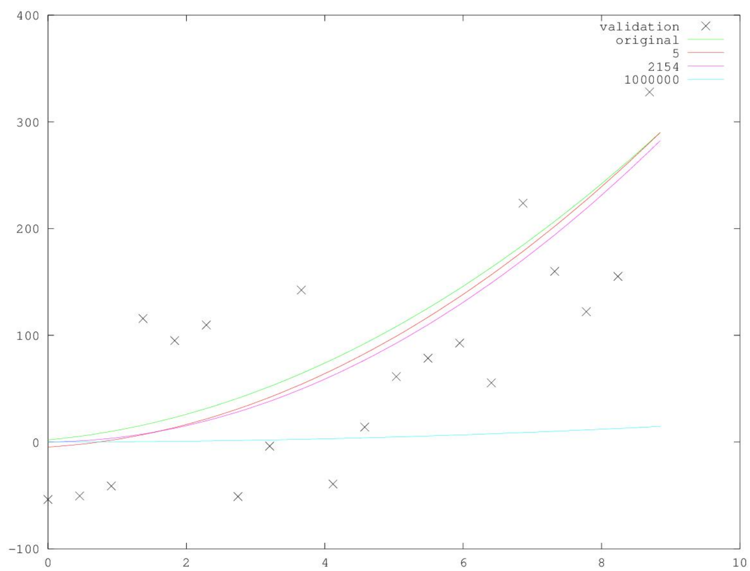


Figure 1: g(x) with 3 approximations (with the noted λ value) and the respective validation set

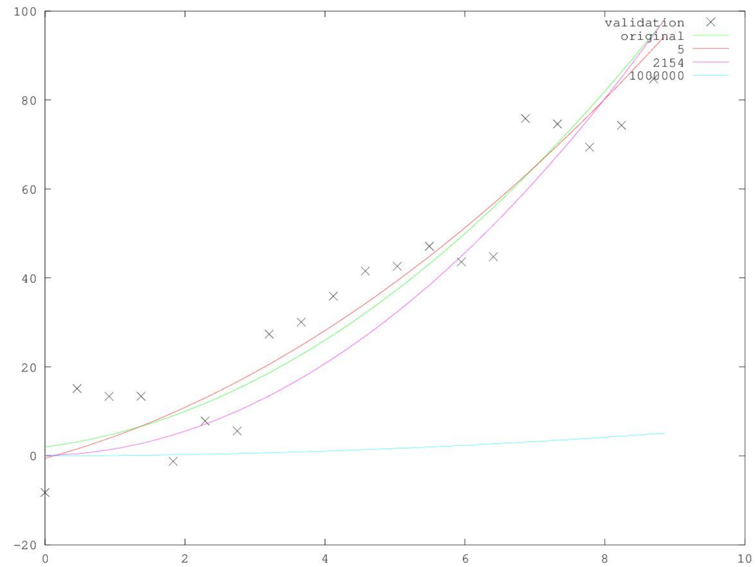


Figure 2: f(x) with 3 approximations (with the noted λ value) and the respective validation set

As it is part of the task, we created a plot with λ on the abscissa and the error on the ordinate as it can be seen in figures 3 and 4. In figures 5 and 6 we can see all the calculated errors for both functions with the best one being indicated.

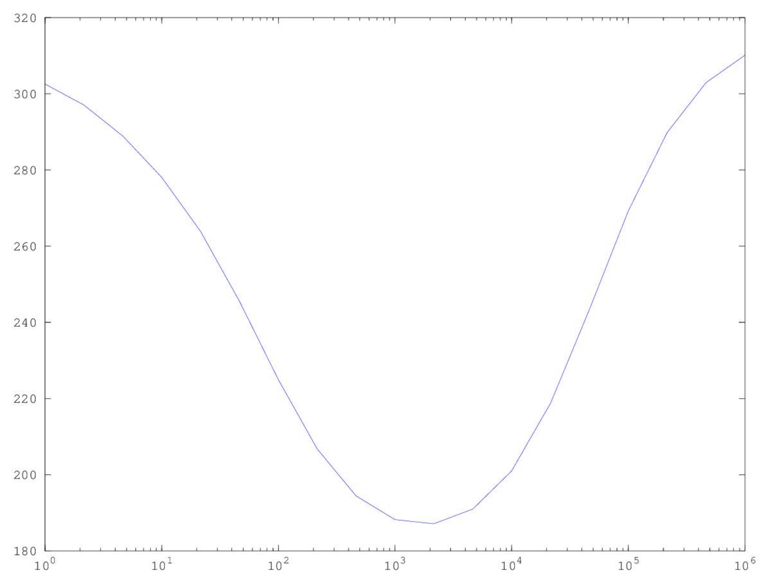


Figure 3: λ and the error plotted on a logarithmic abscissa for $f(x) = 3x^2 + 6x + 2$

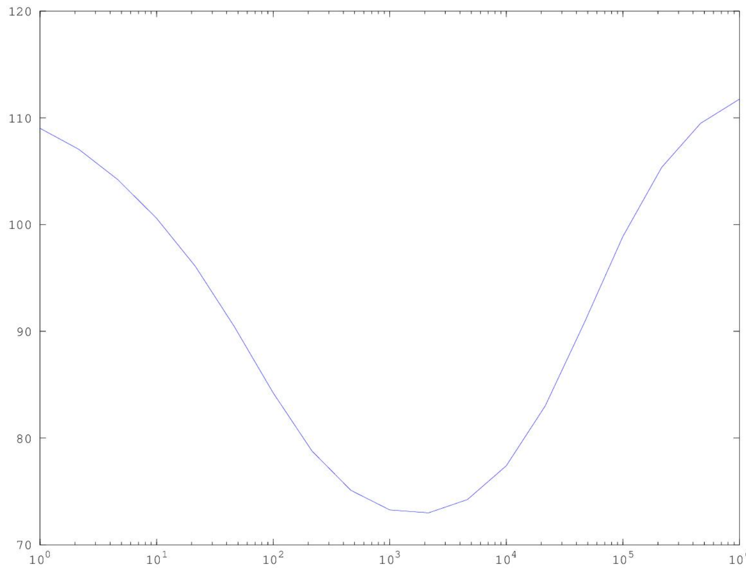


Figure 4: λ and the error plotted on a logarithmic abscissa for $f(x) = x^2 + 2x + 2$

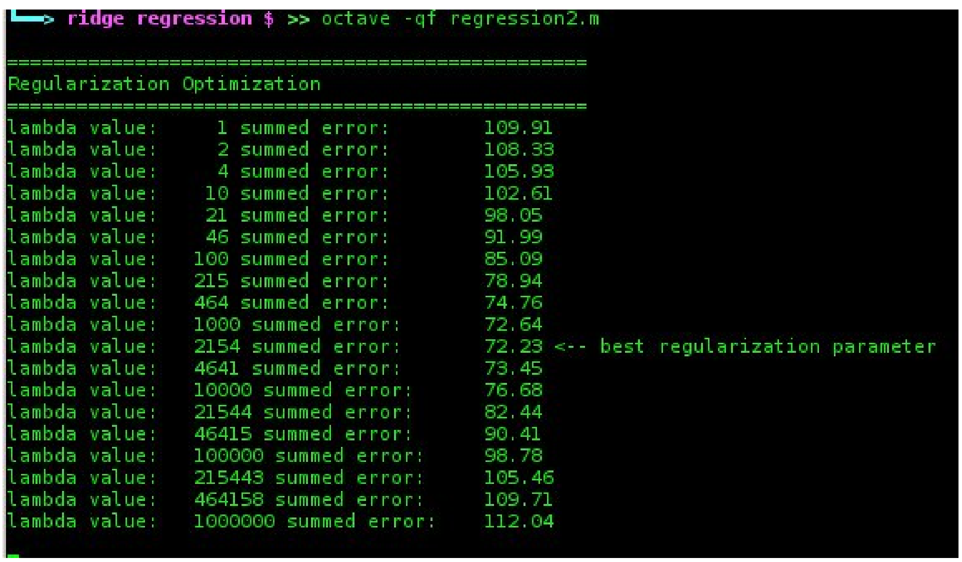
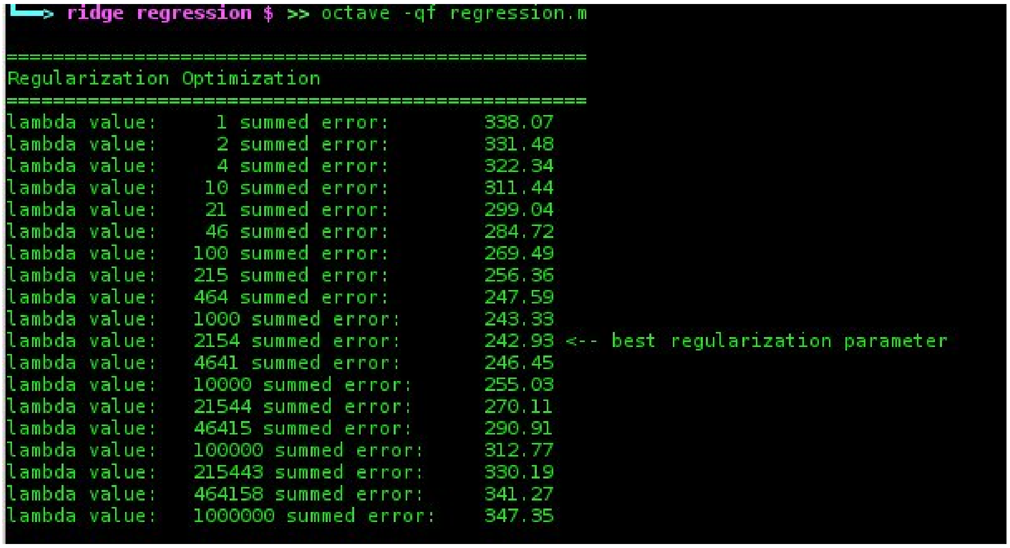


Figure 5: λ and the error for $f(x) = x^2 + 2x + 2$



```
└──> ridge regression $ >> octave -qf regression.m

=================================================
Regularization Optimization
=================================================
lambda value:      1 summed error:        338.07
lambda value:      2 summed error:        331.48
lambda value:      4 summed error:        322.34
lambda value:     10 summed error:        311.44
lambda value:     21 summed error:        299.04
lambda value:     46 summed error:        284.72
lambda value:    100 summed error:        269.49
lambda value:    215 summed error:        256.36
lambda value:    464 summed error:        247.59
lambda value:   1000 summed error:        243.33
lambda value:   2154 summed error:        242.93 <-- best regularization parameter
lambda value:   4641 summed error:        246.45
lambda value:   10000 summed error:       255.03
lambda value:   21544 summed error:       270.11
lambda value:   46415 summed error:       290.91
lambda value:   100000 summed error:      312.77
lambda value:   215443 summed error:      330.19
lambda value:   464158 summed error:      341.27
lambda value:   1000000 summed error:     347.35
```

Figure 6: λ and the error for $g(x) = 3x^2 + 6x + 2$

3 Conclusion

We tested our implementation with different functions without increasing the order of our polynomial as we would have to change the whole implementation of the program. Instead we increased and decreased the weighting factor of each order and our λ values. While doing this we noticed that for the function with the higher slope($g(x) = 3x^2 + 6x + 2$) our approximation with the different λ values is almost identical for small x-values, but with increasing x the regression function will diverge, as seen in figure **??**.
When we use the function with the lower slope($f(x) = x^2 + 2x + 2$) it is the other way around, meaning that the approximation with the different λ values diverge with small x values and converge with bigger x, as seen in figure 2.
As seen in figures 3 and 4 the selection of the regularization parameter λ can have a significant influx on the error of the regression function. Nevertheless this parameter should be chosen carefully because a too big value for λ can increase the error as well (with the limit being a linear regression function with no relation to the original function).