CW 3

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

CO496

Author:
Chengyi Zhang (CID: 01604256)

Date: November 11, 2018

1b

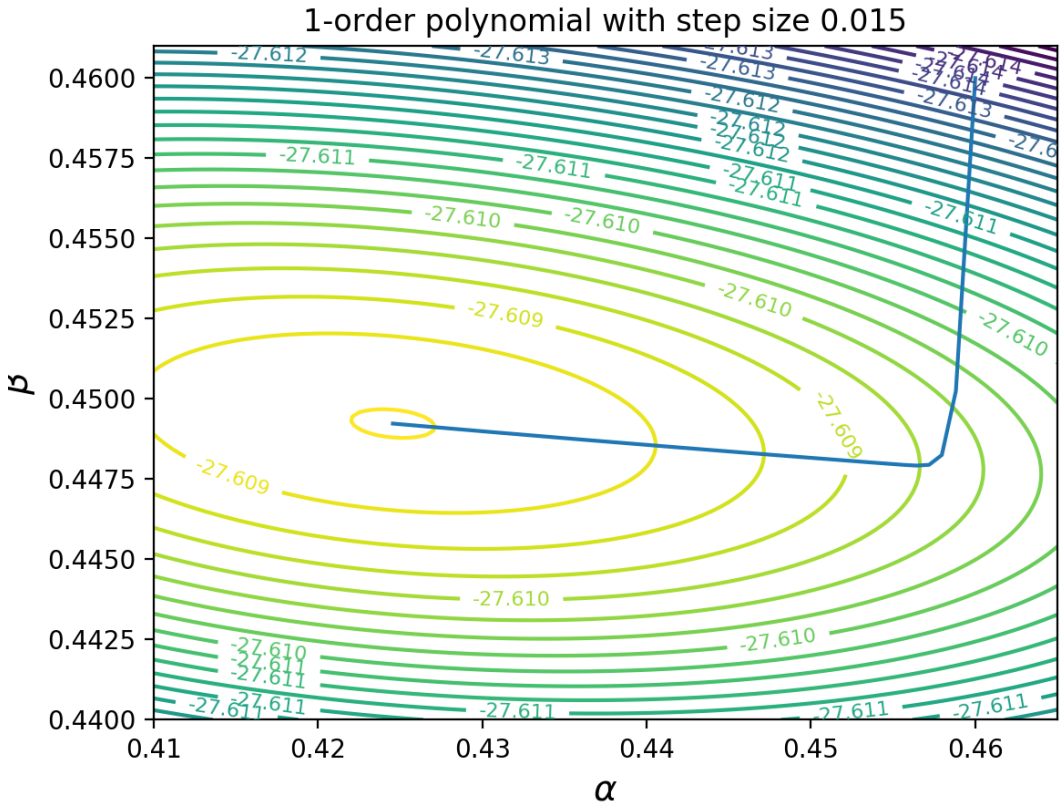

Figure 1: 1-order polynomial with step size 0.013

Starting from initial point $\alpha = 0.46, \beta = 0.46$, after 163 iterations with fix step size 0.015, gradient descent meets first order optimal condition and returns $\alpha = 0.42455483$, *and* $\beta = 0.44923134$.

1c

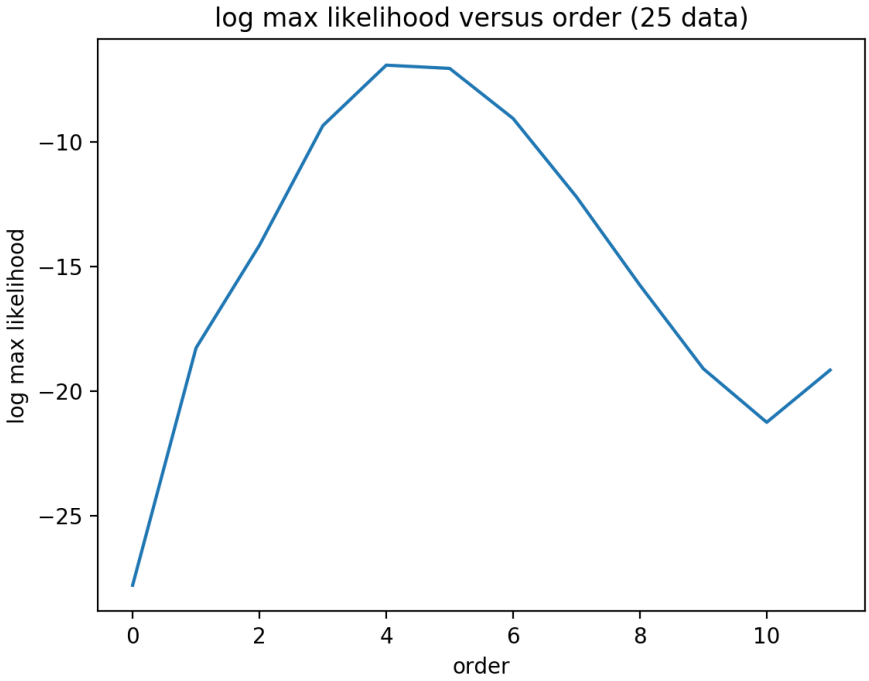

Figure 2: Bayesian Approach

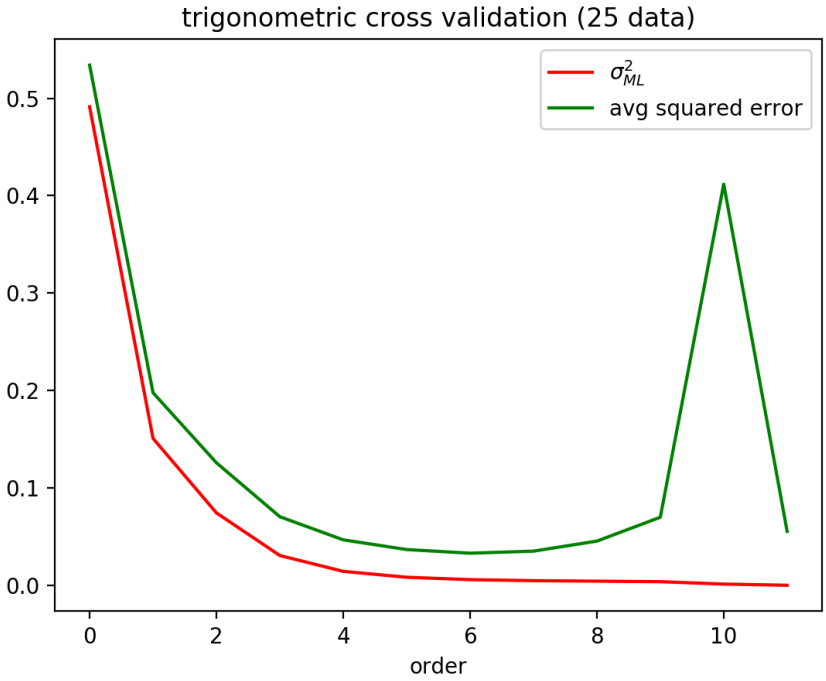

Figure 3: Cross Validation Approach

Two approaches show similar features. In Bayesian Approach, as the order grows, the log maximum likelihood first increases, and then decreases because the model bounds the parameters less strictly. In cross validation, as order grows, the test error decreases and then increases due to over-fitting. Note that in Fig. **??** and Fig. 3, there exists abnormal behaviour when order is eleven. That is because the number of data is too small. When the number of data increases (Fig. 4 and Fig. 5), the behaviour becomes normal.

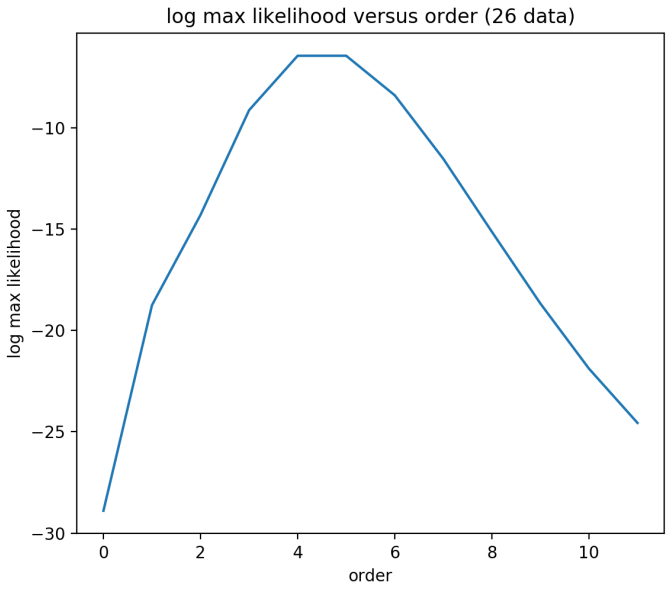

Figure 4: Bayesian Approach with More Data

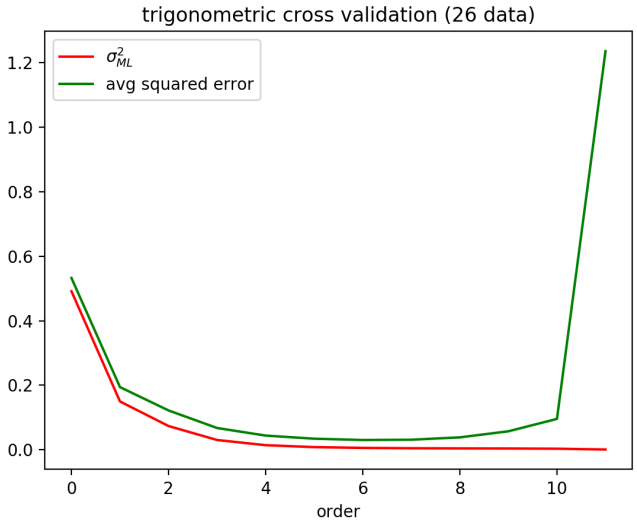

Figure 5: Cross Validation with More Data

1d

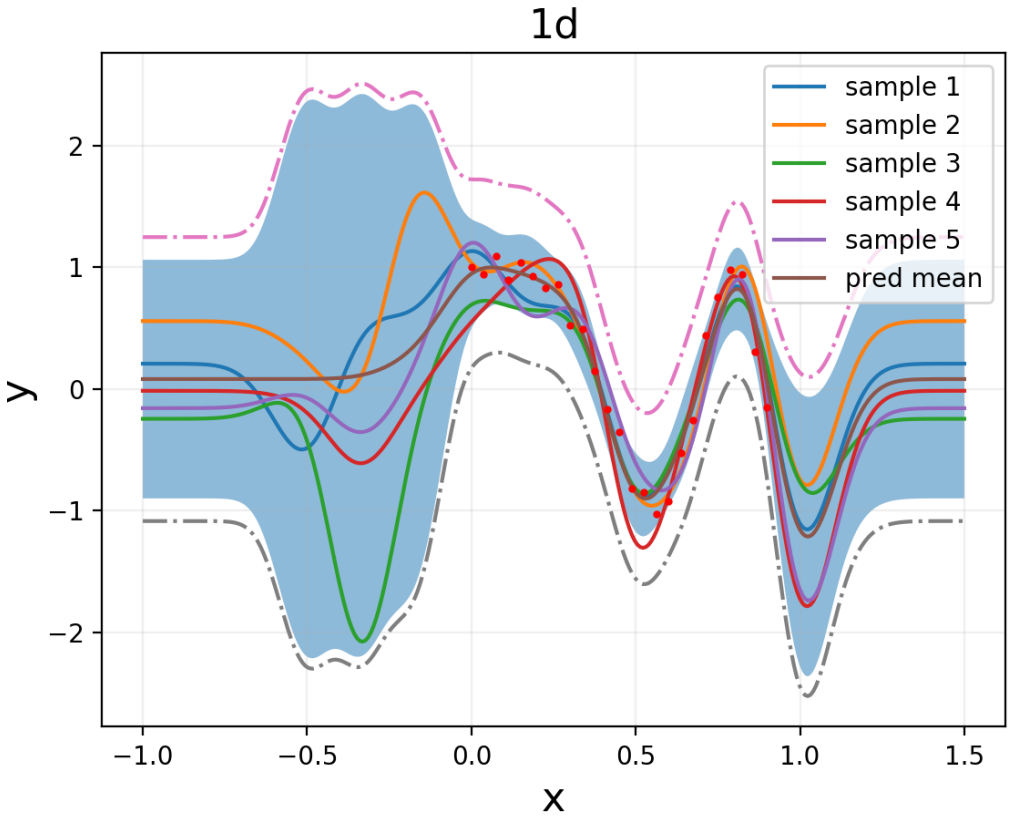

Figure 6: 5 sample ω, predictive mean, and 95% confidence intervals of predictions with(out) noise

The shaded region is a 95% confidence bound where we have 95% confidence that the real $\omega^T \phi(x^*)$ is located inside the shaded regions. And the area bounded by dotted curves include the uncertainty of both posterior and noise, where we have 95% confidence that the real y^* is bounded. From Fig. 6, the model fits well on training interval [0, 0.9]. However, the number of order allows the sampled parameter to vary much and thus the sampled functions are visibly separated. The area of confidence interval gets larger rapidly, which means that the posterior uncertainty is huge.

1e

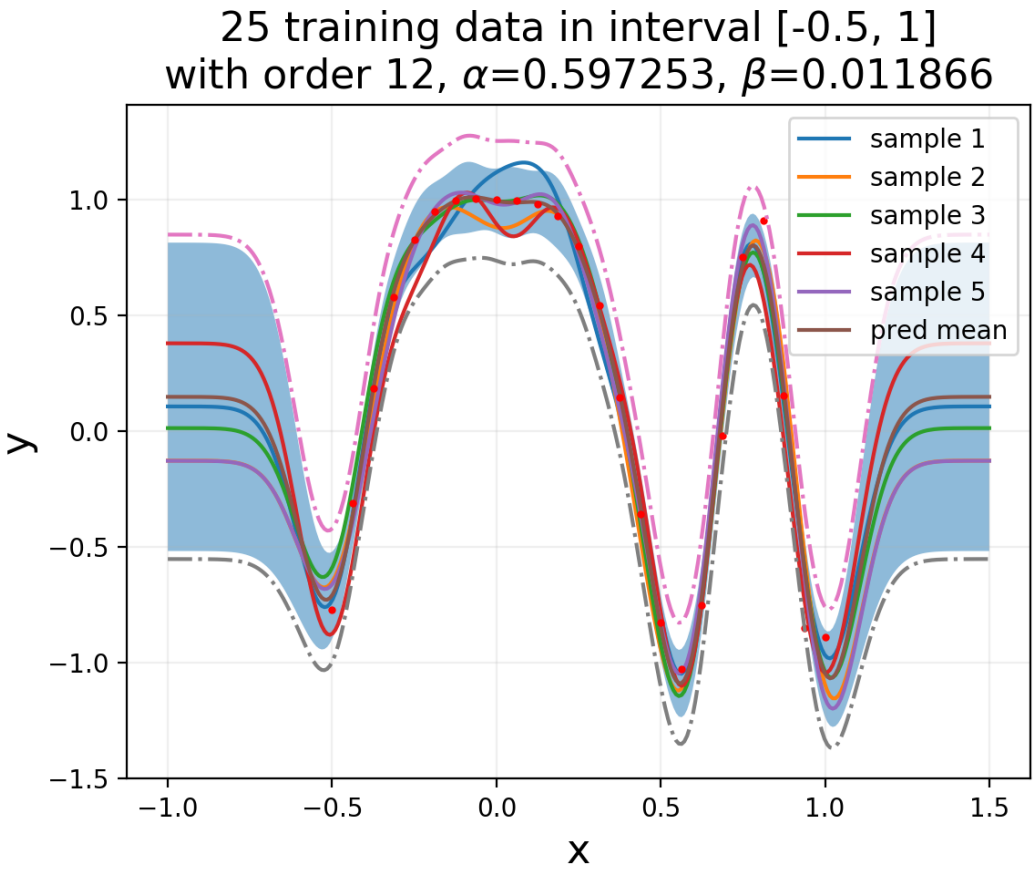

Figure 7: Wider training interval, larger order, and better parameters