Diandra Prioleau

13 September 2018

EEL 5840 – Fundamentals of Machine Learning

HW01

Question 1

The graph in Figure 1 shows the estimated polynomial function when the model order M = 11. The estimated polynomial, which is based on the weight w chosen after fitting the training data to the polynomial, goes through several of the points of the training data. This indicates that overfitting has occurred. Based on the plot in Figure 2, which shows the root-mean-square errors of the training data and test data, the model order M should be in range 5 <= M <= 8 because the test set error is small and will give reasonable representations of the sinc(x) function. At model order M=11, the test set error becomes large, which indicates an over-fitting problem. At this point, the polynomial is more fitted to the training data causing the test set error to increase because it is more difficult to predict values for new data.

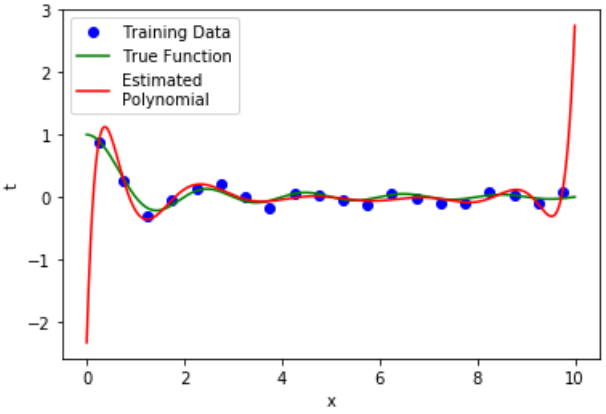

Figure 1 Estimated Polynomial

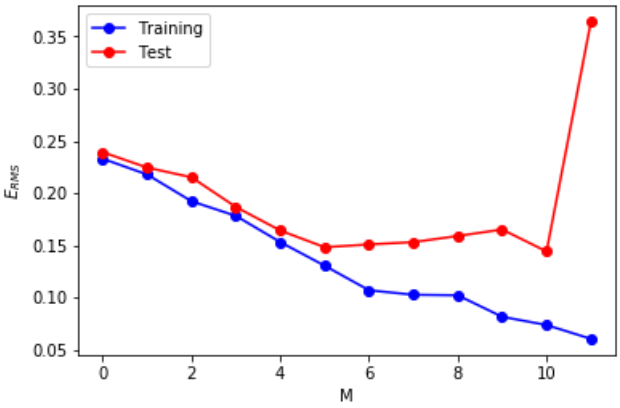

Figure 2 Root-mean-square Errors for Training Data and Test Data

Question 2

Based on Figures 3,4,6-12, it is shown that as the number of samples, N, gets larger, the maximum likelihood solution and the maximum a-posterior solution converge to the same value. When the prior mean is initialized to the wrong value, the mean for the maximum likelihood and the mean for the maximum a-posterior are further apart in the beginning, as shown in Figure 3. In addition, it takes time to overcome the prior in the maximum a-posterior solution.

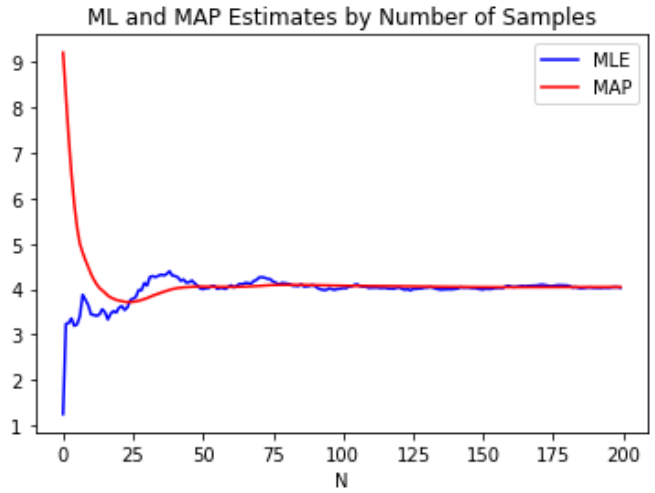

Figure 3 ML and MAP Estimates when Prior Mean is initialized to Wrong Value

When the prior mean is initialized to the correct value, the mean for the maximum likelihood and the mean for the maximum a-posterior are closer in value in the beginning, as shown in Figure 4, than when the prior mean is initialized to the wrong value.

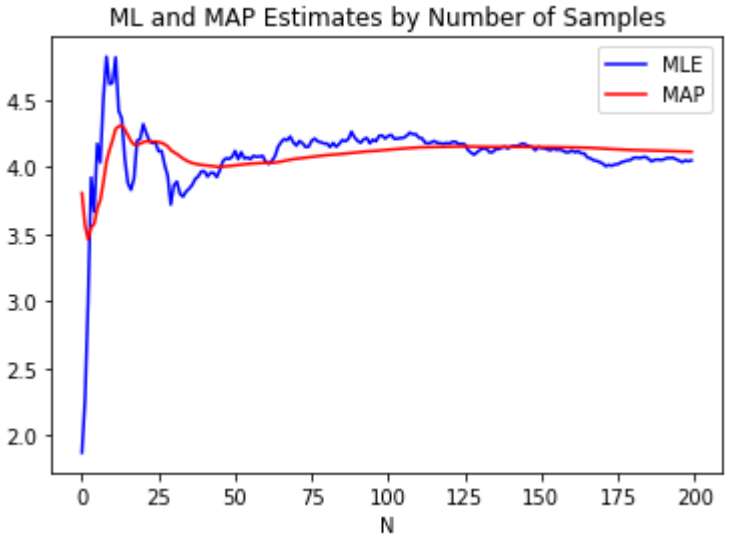

Figure 4 ML and MAP Estimates when Prior Mean is initialized to Correct Value

When the prior variance is small, the mean of the maximum likelihood and the mean for the maximum a-posterior are further apart before converging. If the prior mean was fixed and the prior variance was varying from small to large, the mean of the maximum likelihood and the

mean of the maximum a-posterior would converge. In contrast, if the prior mean was fixed and the prior variance was varying from large to small, the mean of the maximum likelihood and the mean of the maximum a-posterior would diverge as shown in Figure 5. As a result, as shown in Figures 6 -8, when the initial prior variance is smaller, the mean of the maximum likelihood and the mean of the maximum a-posterior are further apart in value in the beginning before converging as the prior mean and the prior variance are updated with the posterior mean and posterior variance; when the initial prior variance is larger, the mean of the maximum likelihood and the mean of the maximum a-posterior are closer in value in the beginning in comparison to when the initial prior variance is smaller. For this problem, as shown in Figures 6-8, although the prior variance is approaching zero, the mean of the maximum likelihood and the maximum a-posterior do not diverge because both the prior mean and prior variance are being updated; the relationship between them causes the mean of the maximum likelihood and the maximum a-posterior to converge to close values.

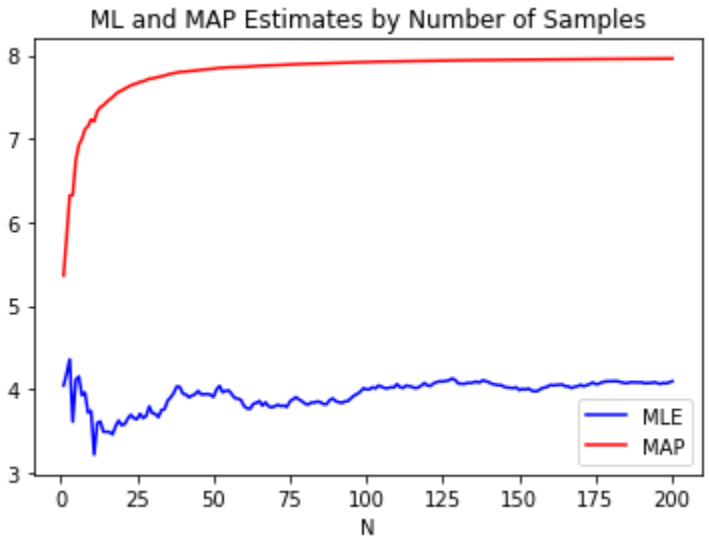

Figure 5 ML & MAP Estimates with Fixed Prior Mean and Varying Prior Variance from Large to Small

Prior variance = 0.5

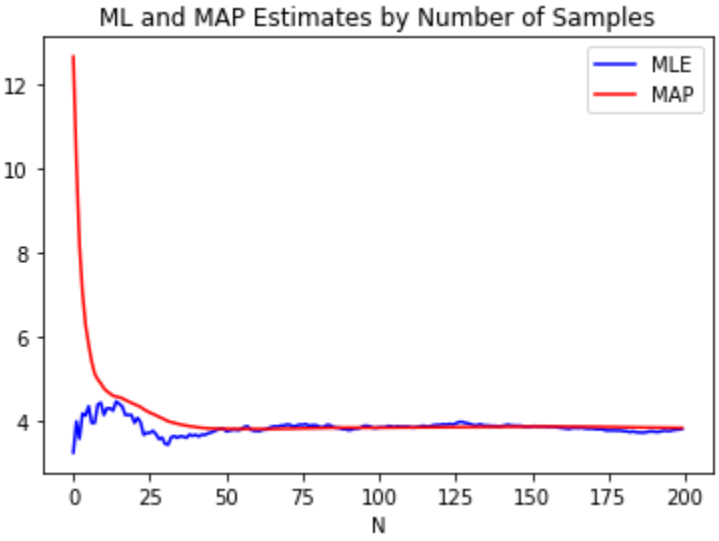

Figure 6 ML & MAP Estimates with Initial Small Prior Variance

Prior variance = 3

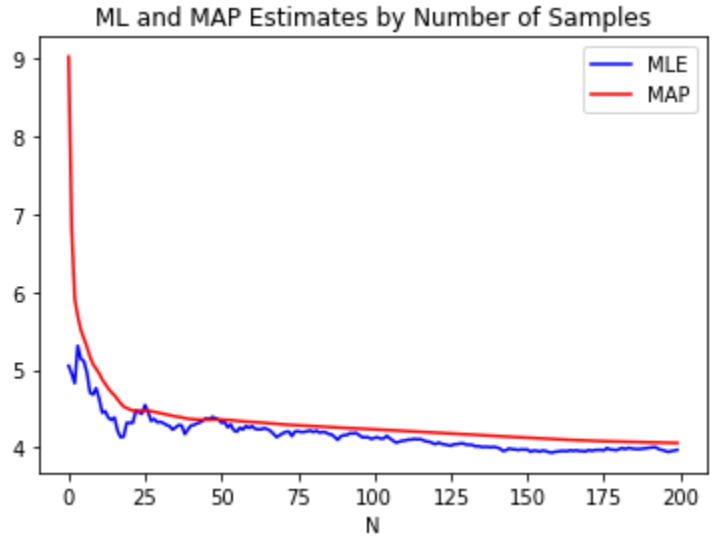

Figure 7 ML & MAP Estimates with Increased Initial Prior Variance

Prior variance = 10

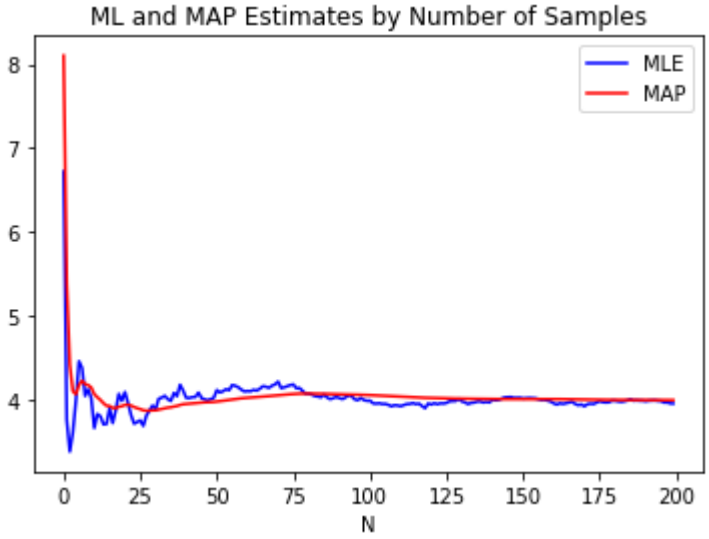

Figure 8 ML & MAP Estimates with Increased Initial Prior Variance

When the likelihood variance is varied from small to large, the initial mean of the maximum likelihood and the maximum a-posterior are closer, as shown in Figures 9-11. For this problem, the maximum likelihood is fixed to the initial value. Therefore, as the number of samples continue to increase, the mean of the maximum likelihood and the mean of the maximum a-posterior continue to converge.

Likelihood variance = 0.5

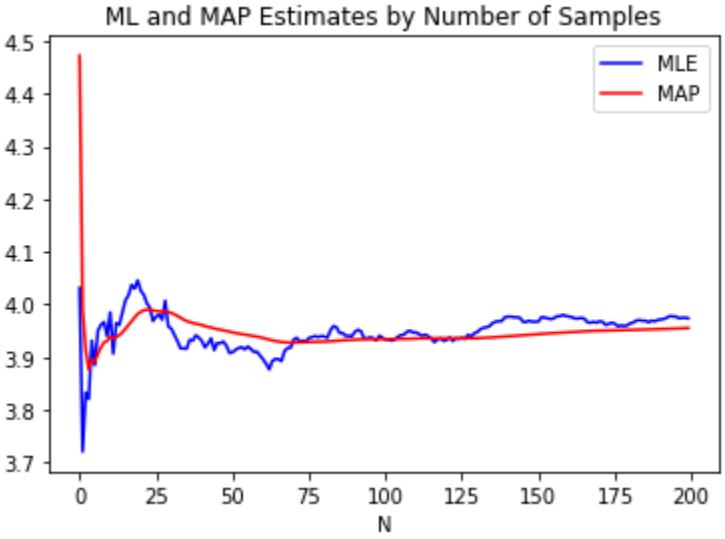

Figure 9 ML & MAP Estimates with Initial Small Likelihood Variance

Likelihood Variance = 2

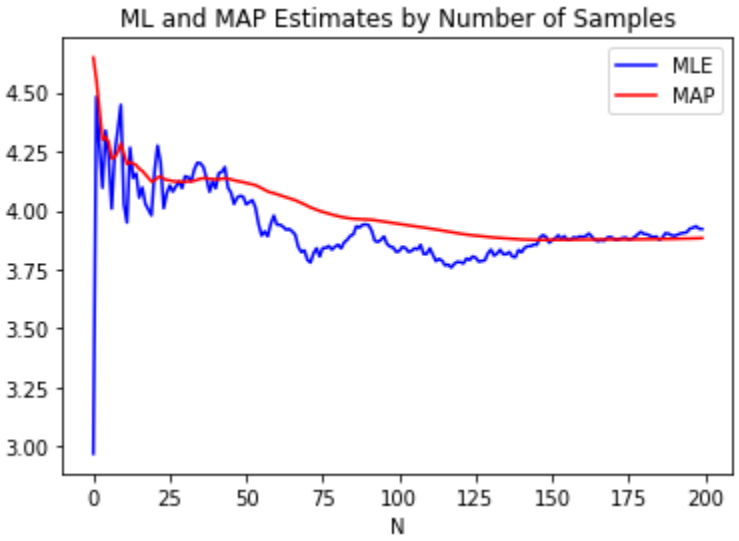

Figure 10 ML & MAP Estimates with Increased Initial Likelihood Variance

Likelihood variance = 4

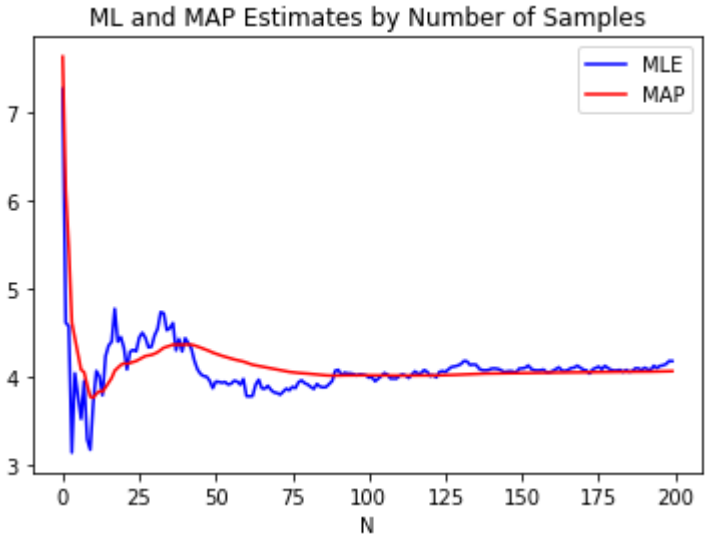

Figure 11 ML & MAP Estimates with Increased Initial Likelihood Variance

The initial values of the prior mean, prior variance, and likelihood variance interact to effect the final estimate of the mean such that their initial values determine how quickly the mean of the maximum likelihood and the mean of the maximum a-posterior will converge. If the initial prior mean is high, it will take more time to overcome the prior in the mean of the maximum a-posterior. In addition, an initial small prior variance will cause the means to be further apart in the beginning before converging; in contrast, an initial larger prior variance will cause the means to be closer in the beginning. An initial small likelihood variance will cause the means to be further apart in the beginning before the maximum

a-posterior converges to the mean of the maximum likelihood; in contrast, an initial large prior variance will cause the means to be closer in the beginning before converging.

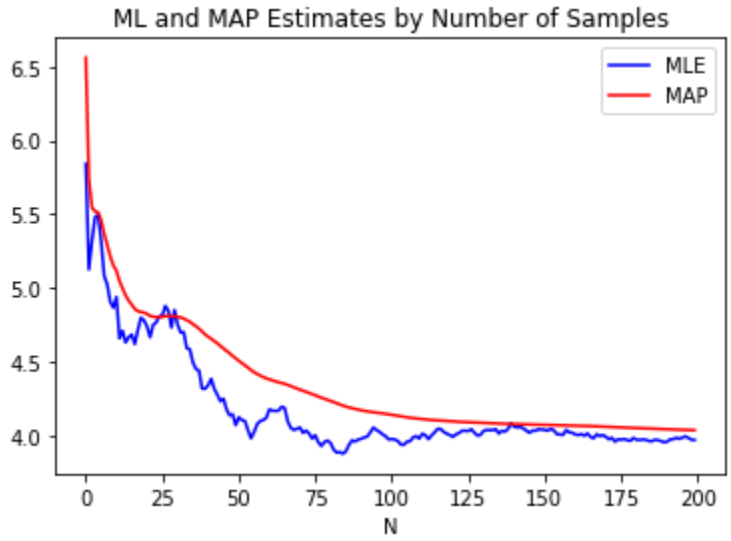

Figure 12 ML & MAP Estimates as Number of Samples Increase

References

1. Textbook – Pattern Recognition & Machine Learning by Christopher M. Bishop
2. Online Resource:
    a. https://atmos.washington.edu/~dennis/MatrixCalculus.pdf
    b. https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf