Student Name: Irene Chang
Collaboration Statement:

Turning in this assignment indicates you have abided by the course Collaboration Policy:

`www.cs.tufts.edu/comp/136/2022s/index.html#collaboration-policy`

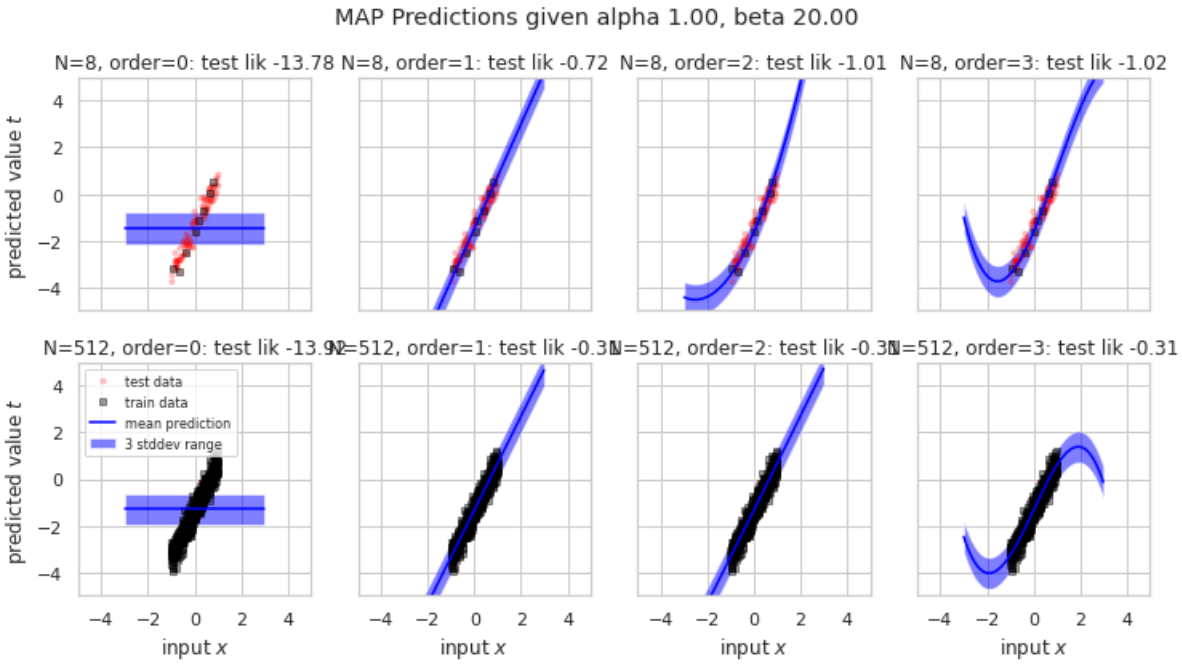
Total hours spent: 10 hours

I consulted the following resources:

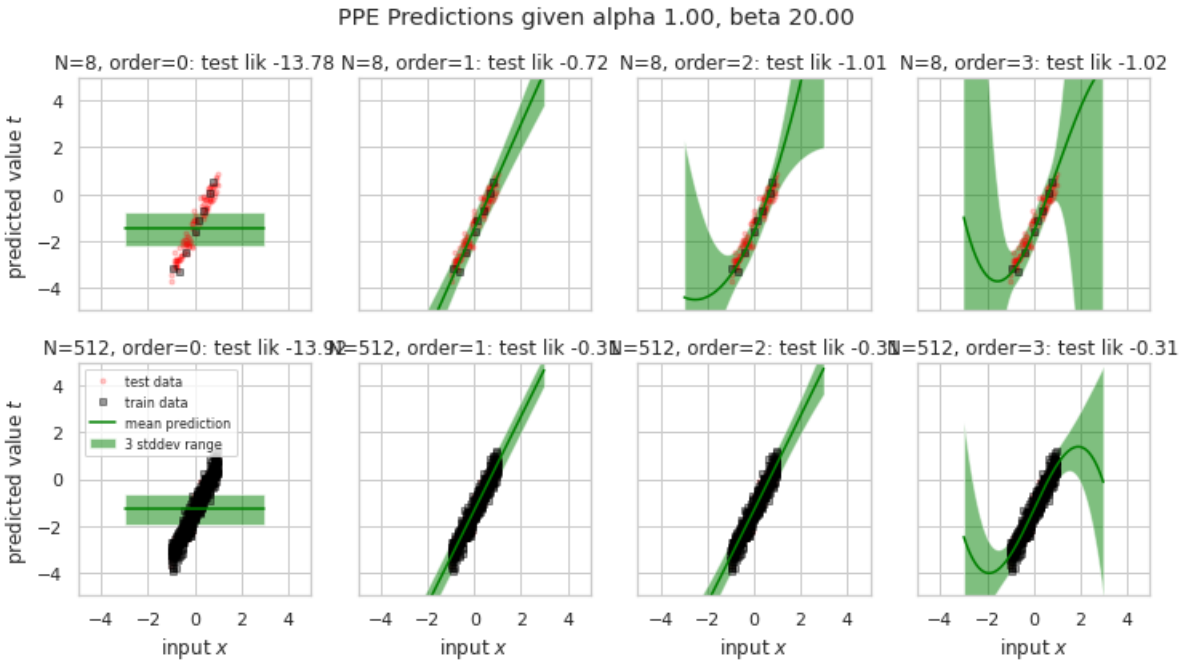
- TA Kapil, Professor Lage on Piazza
- textbook

Problem 1

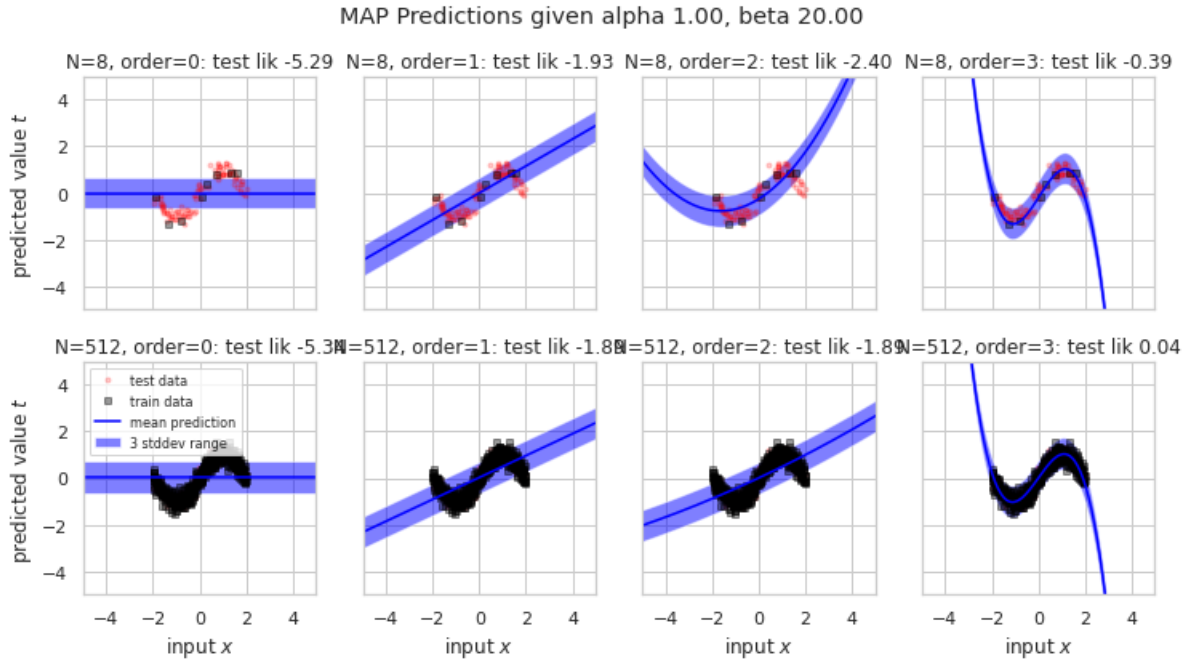
1a.



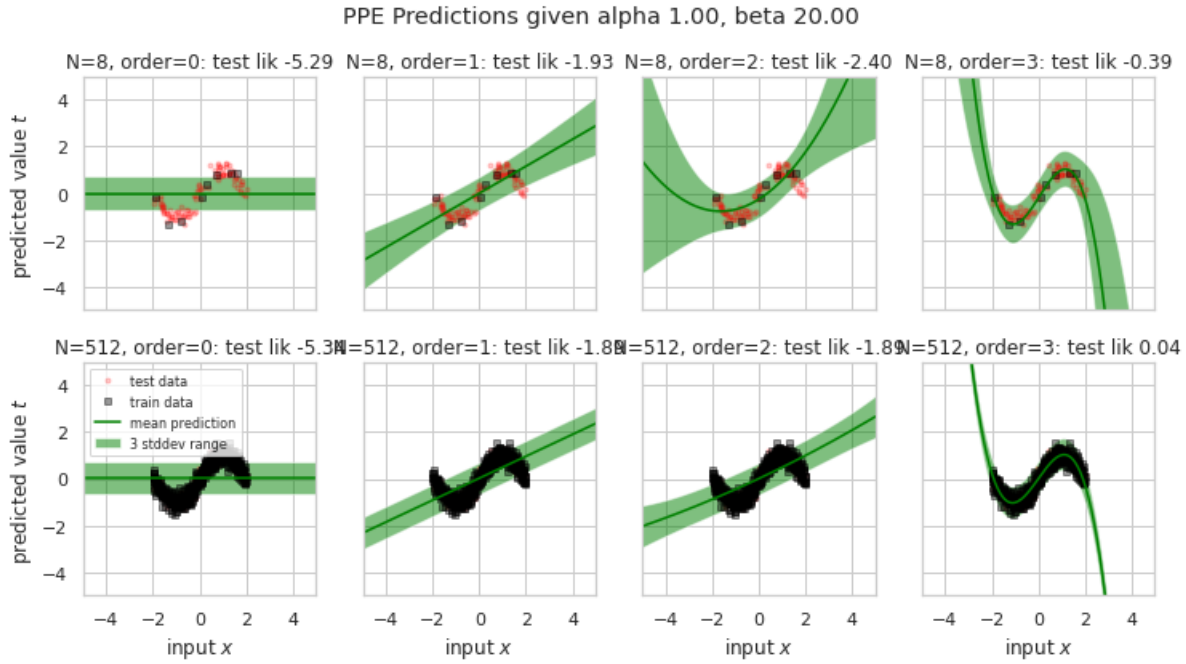
1b.



1c.


MAP Predictions given alpha 1.00, beta 20.00
N=8, order=0: test lik -5.29
N=8, order=1: test lik -1.93
N=8, order=2: test lik -2.40
N=8, order=3: test lik -0.39
N=512, order=0: test lik -5.3
N=512, order=1: test lik -1.8
N=512, order=2: test lik -1.89
N=512, order=3: test lik 0.04
test data
train data
mean prediction
3 stddev range
predicted value t
input x

1d.


PPE Predictions given alpha 1.00, beta 20.00
N=8, order=0: test lik -5.29
N=8, order=1: test lik -1.93
N=8, order=2: test lik -2.40
N=8, order=3: test lik -0.39
N=512, order=0: test lik -5.3
N=512, order=1: test lik -1.8
N=512, order=2: test lik -1.89
N=512, order=3: test lik 0.04
test data
train data
mean prediction
3 stddev range
predicted value t
input x

1e. Think about the heldout predictions of these estimators of new outputs given new inputs (e.g. as in plots 1a and 1c). How are the PPE mean estimates different from the MAP mean estimates? How are the PPE variance estimates different from the MAP variance estimates? Base your answer on what you expect from the underlying math, as well as any plots you have made.

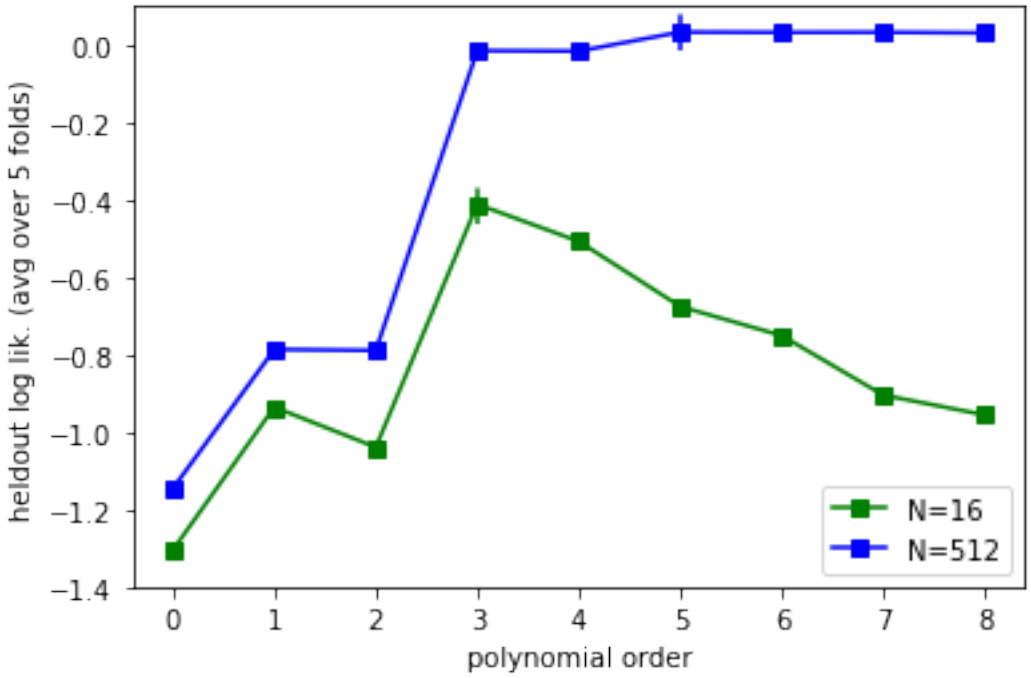
The likelihood of both methods follow the normal distribution, whose means follow a very similar formula of:

$$(\text{const} \times \Phi^T \Phi + \text{const} \times I)^{-1} \Phi^T t_N$$

Thus we can observe in the graph the mean prediction for both methods follow almost identical curves. The difference between these two methods is in the standard deviation (and hence variance) around the mean curves. We observed that for both datasets, PPE yields an equal or bigger standard deviation range than MAP. This is because the normal distribution of the likelihood of MAP estimator has a constant variance of β^{-1}, while that of PPE depends on the input test data ($\sigma_N^2(x^*)$). Also observe that this discrepancy in the variance/standard deviation is more significant for smaller dataset (the plots for sample size of 8 in this case saw a much bigger difference between the two methods). As the estimator receives more data, the variance shrinks, which explains why the stdev ranges for N = 512 for both methods are very similar to each other.

Problem 2

2a.

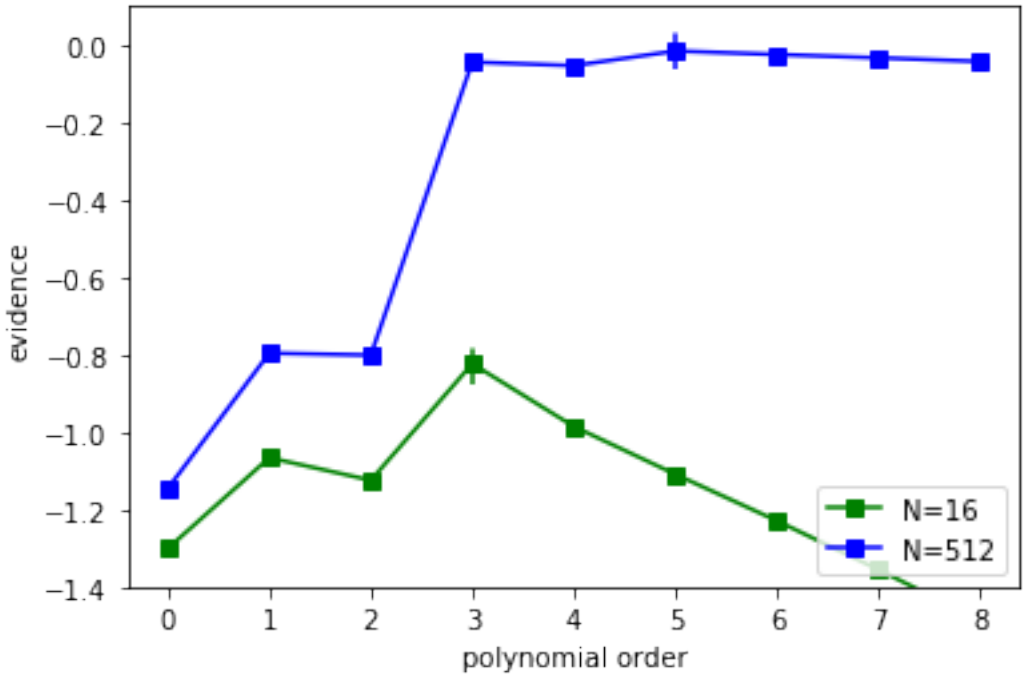


2b. Report verbatim the selected hyperparameter values as well as the model's score on the test set

order = 5
alpha = 0.0001
beta = 15.8
test likelihood score: 0.0001655
required time = 7.62 sec

Problem 3

3a.



3b. Best hyperparameters for N=512 training set Report verbatim the selected hyperparameter values as well as the model's score on the test set. Like in 2b

order = 5
alpha = 1
beta = 15.8
test likelihood score: 0.0001631
required time = 0.67 sec

3c. Which grid search was more expensive in terms of run time, 5-fold CV in problem 2 or the evidence in problem 3?

5-fold grid search CV in problem 2 is more expensive in terms of runtime. (The required time for the best model in q3 is 0.67s, compared to 7.62s in q2)

3d. Did the final model chosen in Problem 2 grid search differ substantially from the best model from the Problem 3 grid search? Discuss any differences in selected hyperparameter values as well as test set performance.

The model chosen in part 2 differs from the model chosen in part 3 in the parameter alpha. We got the same order (5) and beta value (15.8). The test set log likelihood scores are almost identical and are both high compared to other models of different parameters (0.0001631 in part 3 and 0.0001655 in part 2). We also see that the score curves for both parameter selection methods, for N = 512, are very similar. So, without taking runtime cost into account, both models perform equally well. With runtime being considered, the model in 3 would be a better option.