

Student Name: Guangyu Lin

Collaboration Statement:

Total hours spent: a week

I discussed ideas with these individuals:

- TODO
- TODO
- ...

I consulted the following resources:

- office hour with Professor and TA
- textbook
- ...

By submitting this assignment, I affirm this is my own original work that abides by the course collaboration policy.

1a

The score function is :

$$\frac{\sum_{n=1}^N \log \text{NormPDF}(t_n | w_{map}^T \phi(x_n), \beta^{-1})}{N} \quad (1)$$

1b

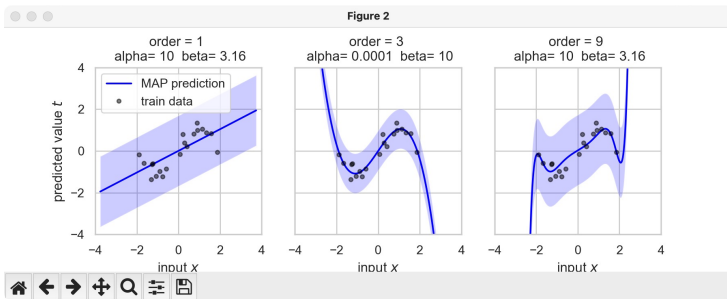
By looking at the score on different beta and different order, we can see when order = 1, it preferred beta = 4, the best score is -0.7378648071971698. When order = 3, it preferred beta = 100 the best score is 1.0379688636129498. When order = 9, it preferred beta = 2500, the best score is 2.993073441655005. When we are using MAP estimator, the variance is a fixed parameter which is the inverse of beta. So, when beta becomes bigger, the distribution becomes more concentrate. When beta becomes smaller, the distribution becomes more spread out. And when order is small, the gap between estimated mu and true mu is big and the estimated mu will be more likely to lie on the distribution with bigger variance which has smaller beta. When order is big, the gap will increasing and more likely on the more concentrated distribution which has bigger beta. Therefore, the small order will prefer small beta and

large order will prefer large beta.

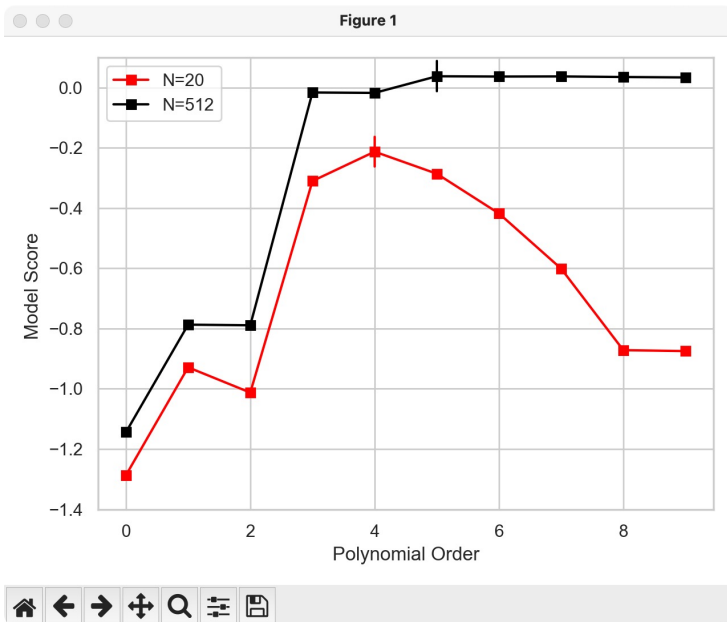
1c

The reason is PPE can help us determine whether our model is good for the test data. By looking at the visuals of PPE and MAP, we can find PPE has variance outside of the predictions that far from training data. But MAP didn't show up any variance besides the training data points. Therefore, MAP is hard to tell us the how bad the model is with these data that far from training data but PPE does.

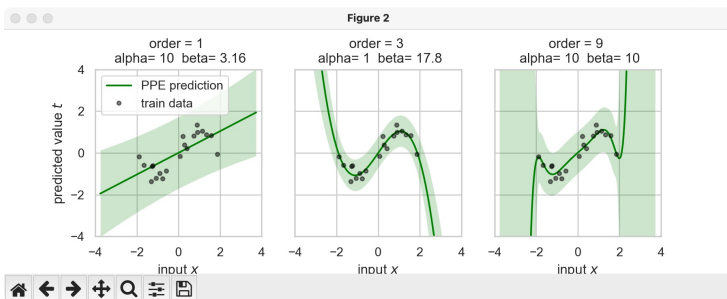
2a



2b



3a



3b

