Homework 1

Discussion of Question 1:

The code fits sinc(x) function with high polynomial model. The polynomial model order varies from 1 to 13. The curve of root-mean-square error to model order is shown as Fig. 1.1.

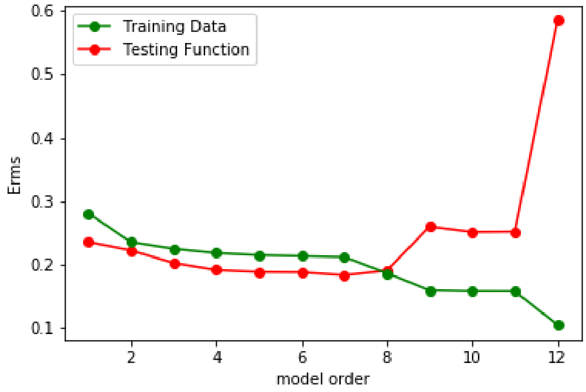

Fig. 1.1 The curve of root-mean-square error to model order

Fig. 1.1 demonstrates that the root-mean-square of training data points and estimated polynomial, which is presented by the red curve, decreases while the model order grows. This also happens on the green curve, which is determined by values of training data points and estimated polynomial. When model order is 9, the red curve turns to rise. This illustrates that when model order is 9, the fitting result overfits the training data. Fig. 1.2 and Fig. 1.3 prove this conclusion as well. Thus, the maximum model order, M, should be selected as 8 to avoid over-training.

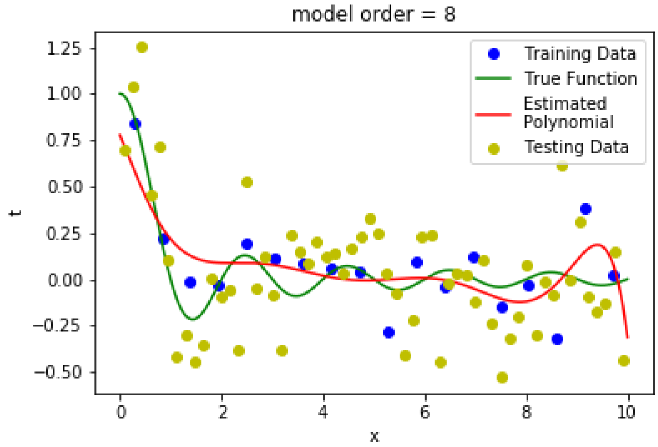

Fig. 1.2 The curve of true function and estimated polynomial when model order is 8

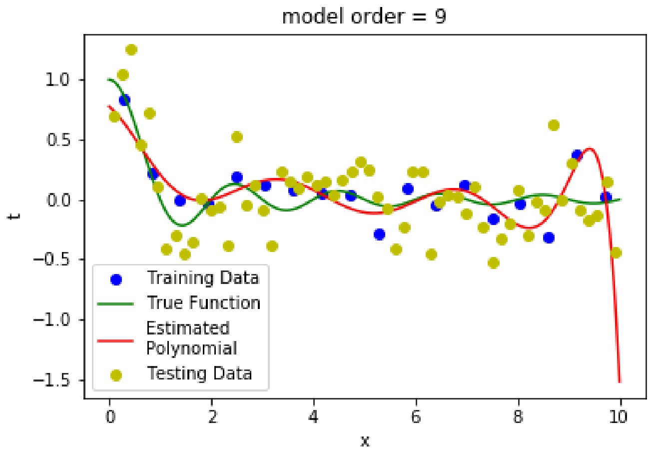

Fig. 1.3 The curve of true function and estimated polynomial when model order is 9

The curve with model order from 1 to 13 could be generated when the code for Question 1 is executed.

Discussion of Question 2:

The code generates a series of data points subjected to Gaussian distribution, whose prior mean is 4 and prior variance is 0.2. Then it computes the maximum likelihood estimation (MLE) of the data points. It also computes maximum a posteriori (MAP) with prior estimation, and prior estimation is the main difference between MLE and MAP. MAP contains prior estimation, which could affect the values of MAP solution of mean. However, when the number of data points is large enough, it could gradually ignore the effect of prior estimation and finally get an accurate estimation.

For the question,

what happens when the prior mean is initialized to the wrong value? To the correct value?

The prior mean is initialized to 8, while the true mean is 4. The behavior of MLE and MAP is presented in Fig. 2.1.

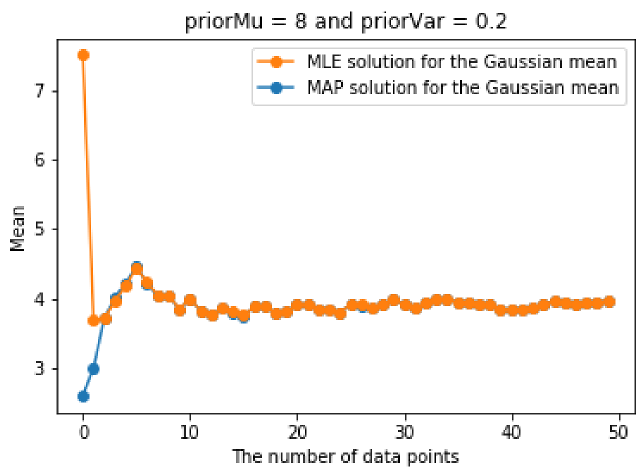

Fig. 2.1 The curve of Mean to number of data points with prior mean equals 8

Fig. 2.1 demonstrates that even though the prior mean is totally wrong. The values of MLE and MAP solution of the Gaussian mean converge to the true value with the growing number of data points.

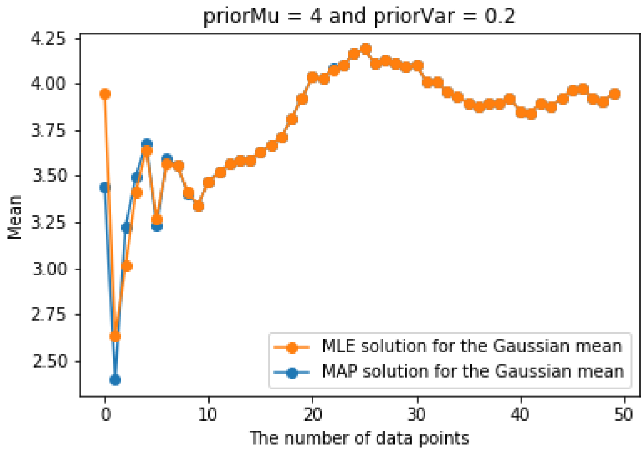

Fig. 2.2 The curve of Mean to number of data points with prior mean equals 4

Fig. 2.2 presents that when the prior mean equals true value, the curve fluctuates first and gradually settle down to the true value. It acts as the prior mean is wrong.

What happens as you vary the prior variance from small to large?

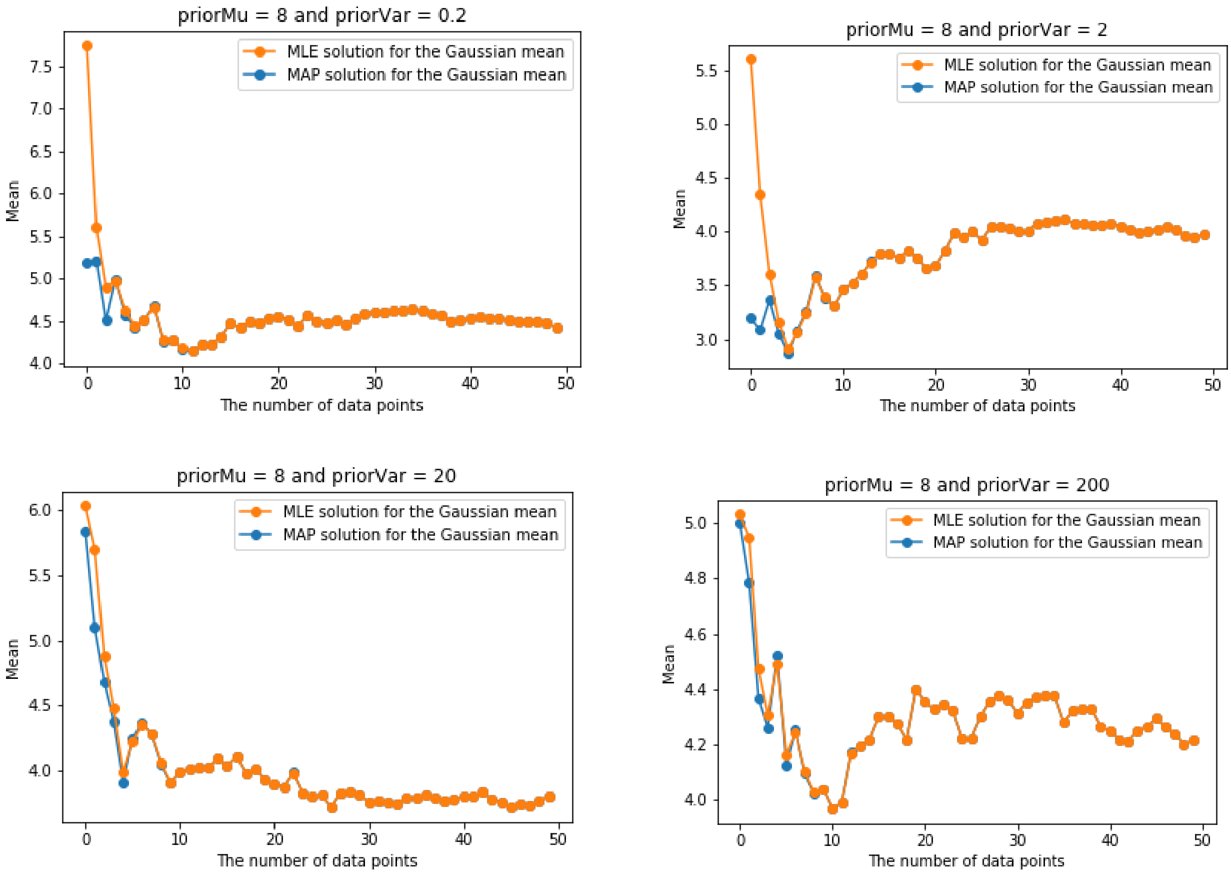

Fig. 2.3 The curve of Mean to number of data points with different prior variance

The real variance is set to 2. Fig. 2.3 shows that the curves converge to the true value with the growing number of data points with different prior variances, and different prior variances don't actually affect the speed of convergence and the amplitude of curve fluctuation a lot.

What happens when the likelihood variance is varied from small to large?

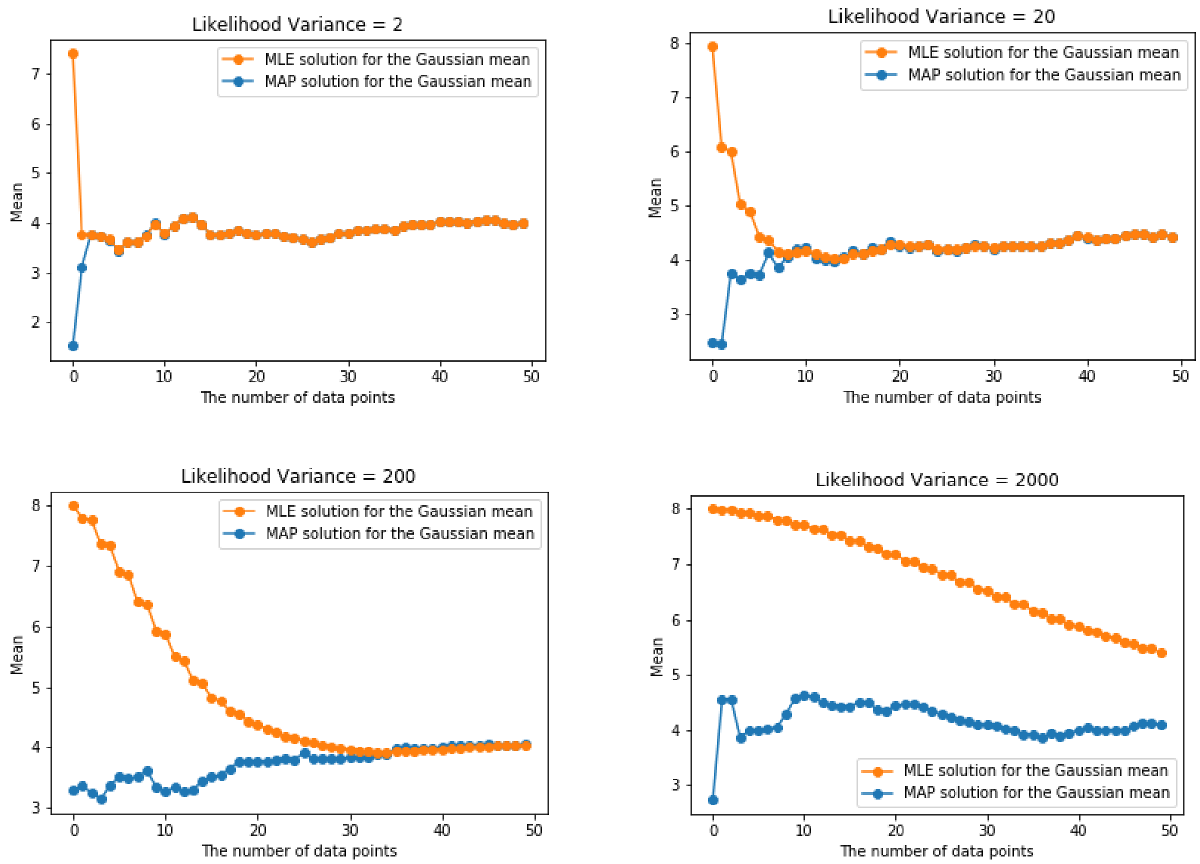

Fig. 2.4 The curve of Mean to number of data points with different likelihood variance

Fig. 2.4 shows that the speed of convergence becomes slower with growing likelihood variance, in other words, more data points are needed for MLE and MAP solutions when the variance is larger, and the curve struggles against more fierce fluctuation as well.

How do the initial values of the prior mean, prior variance, and likelihood variance interact to effect the final estimate of the mean?

From the figures above, it could be concluded that prior mean, prior variance and likelihood variance don't effect the final estimate of the mean with enough data points. With a certain number of data points, prior mean and prior variance cannot effect the final estimate, but likelihood variance is able to effect it if likelihood variance is large enough.

Discussion of Question 3:

In [1], the authors present matrix derivative formula expressions, which is available for Question 3 to Question 6.

$$f(x) = 3x^T x + 4y^T x - 1$$

$$\frac{\partial f}{\partial x} = 3(x^T \frac{\partial x}{\partial x} + x^T \frac{\partial x}{\partial x}) + 4y^T$$

$$= 3 \cdot 2x^T + 4y^T$$

$$= 6x^T + 4y^T$$

Discussion of Question 4:

$$f(x) = -10x^T Qx + 4y^T x + 2$$

$$\frac{\partial f}{\partial x} = -10(\frac{\partial x^T}{\partial x} Qx + \frac{\partial (Qx)^T}{\partial x} x) + 4y^T$$

$$= -10(Q + Q^T)x + 4y^T$$

$$= -20Qx + 4y^T$$

Discussion of Question 5:

$$f(x) = 8x^T Qx - 2y^T Q^T x + 6$$

$$\frac{\partial f}{\partial x^T} = 8(\frac{\partial (x^T Qx)^T}{\partial x})^T - 2(\frac{\partial (y^T Q^T x)^T}{\partial x})^T$$

$$= 8(\frac{\partial (x^T Q^T x)}{\partial x})^T - 2(\frac{\partial (x^T Qy)}{\partial x})^T$$

$$= 8(Q + Q^T)x - 2(Qy)^T$$

$$= 16Qx - 2y^T Q^T$$

$$\frac{\partial f}{\partial Q} = 8xx^T - 2yx^T$$

Discussion of Question 6:

$$f(x) = \|4x\|_2^2$$

$$\frac{\partial f}{\partial x} = \frac{\partial (16x^T x)}{\partial x}$$

$$= 32x^T$$

[1] Petersen K B, Pedersen M S. The matrix cookbook[J]. Technical University of Denmark, 2008, 7(15): 510.