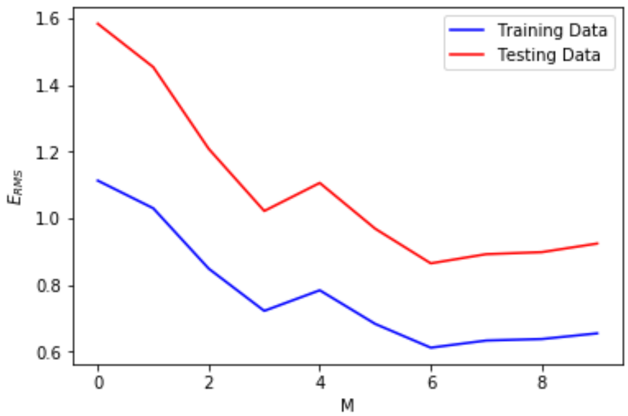
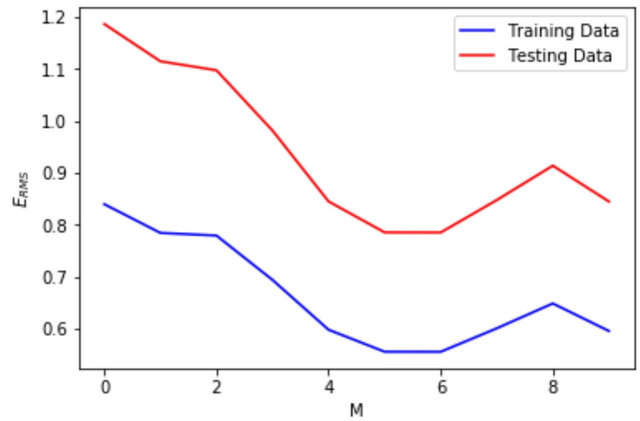
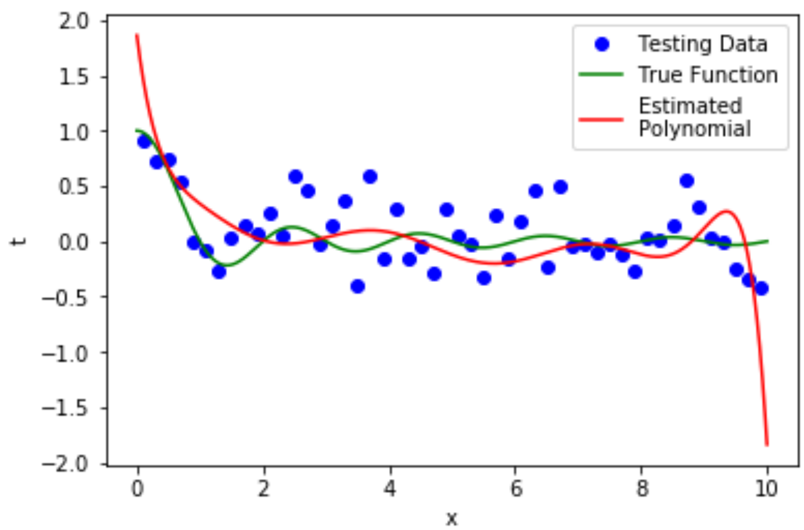
1.





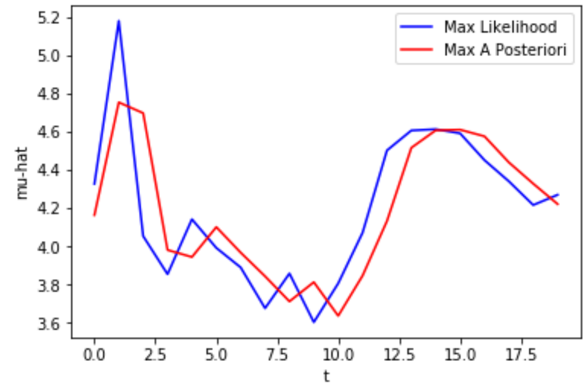
The charts above were generated by my Python code for question 1 in 2 different runs of the code. From many such iterations of my script, it looks like M=5 or M=6 are the best order to fit the polynomial, but there is a lot of variation among separate iterations.



This plot of the estimated polynomial of M=6 alongside the true function and also the testing data shows that this does a relatively decent job of predicting the testing data but generally looks overfitted on the left and the right, something that carries over in many other iterations I've run of M=5 and M=6.

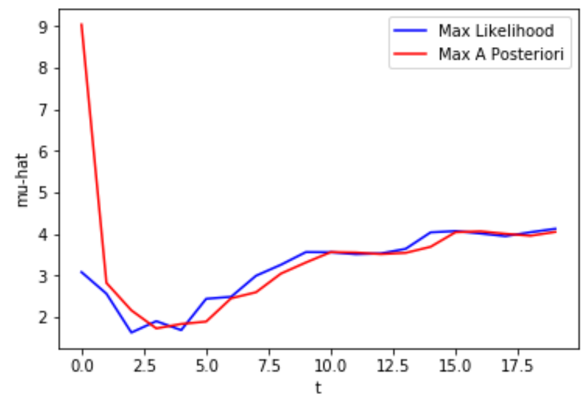
2. The following charts are a graph of the MLE and MAP plotted at the t^{th} iteration.

Case 1:	
trueMu	4
trueVar	2
priorMu	4
priorVar	2



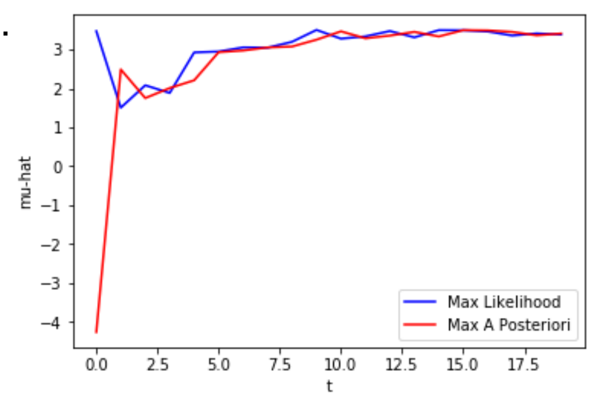
Case 1 just kind of set a barometer for me going into this analysis. Since the prior are set close to the true parameters, the outcome of the MLE and MAP are similar. As you can see, the MAP is less affected by spikes resulting from sample variation, and since it takes the previous iteration's posterior as its prior, the spikes are shifted later from those in the MLE.

Case 2:	
trueMu	4
trueVar	2
priorMu	15
priorVar	0.005



Case 2 shows what happens when the prior mean is set higher than the true mean. It is evident that the MAP initially is more resistant to change from the prior belief and takes a few iterations to converge to the MLE.

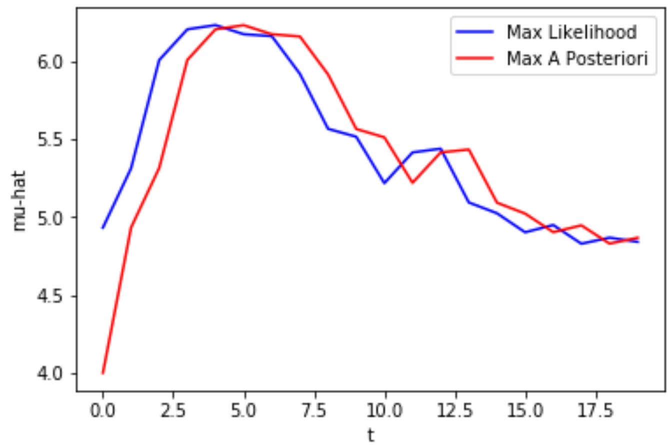
Case 3:	
trueMu	4
trueVar	2
priorMu	-12
priorVar	2



Case 3 shows what happens when the prior mean is set lower than the true mean. This case is very similar to what happened in case 2.

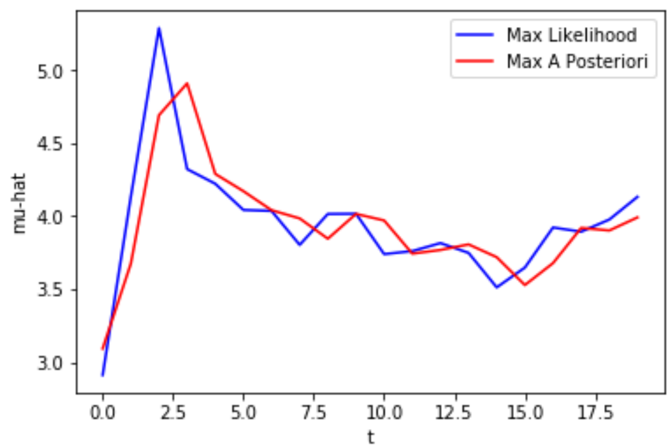
Case 4 illustrates what happens when the prior variance is lower than the true variance: the max a posteriori becomes effectively a shifted version of the MLE of the iteration before.

Case 4:	
trueMu	4
trueVar	2
priorMu	4
priorVar	0.025



Case 5 illustrates what happens when the prior variance is larger than the actual variance, bringing the max a posteriori closer to a flatter, less jagged version of the MLE of the current iteration.

Case 5:	
trueMu	4
trueVar	2
priorMu	4
priorVar	10



In summary, when our prior mean is initialized to the wrong value, the maximum a posteriori takes longer to converge to the true value of the mean, whereas when initialized to the right value converges to the true mean faster and with less susceptibility to bias from sampling variation. As the prior variance goes from small to large, the MAP converges closer to the MLE. When the likelihood variance is varied from small to large, it weights the MAP more toward the prior term.

The initial value of the prior mean gives the MAP a starting point to protect it from outliers, from which it can either stay around there if the prior mean is correct or go toward the true mean if the prior mean is incorrect.

The initial value of the prior variance allows you to control the weighting of the prior belief, or or “how strongly you impose the prior”: the smaller the variance of the prior, the more important the prior term, whereas the smaller the variance of the likelihood, the more important the likelihood estimate.

3.

3. $f(x) = 3x^Tx + 4y^Tx - 1, \quad x, y \in \mathbb{R}^d$

$$= 3(x_1 \cdots x_d)\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} + 4(y_1 \cdots y_d)\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} - 1$$

$$= 3[x_1^2 + \cdots + x_d^2] + 4(x_1y_1 + \cdots + x_dy_d) - 1$$

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \cdots \ \frac{\partial f}{\partial x_d}\right]^T = [6x_1 + 4y_1 \ \cdots \ 6x_d + 4y_d]^T$$

$$= (\underbrace{6x^T}_{1\times d} + 4y^T)^T = \underbrace{6x}_{d\times d} + \underbrace{4y}_{d\times 1}$$

4.

Note: Q is symmetric $\Rightarrow q_{ij} = q_{ji} \ \forall i,j$

4. $f(x) = -10(x_1 \cdots x_d)\begin{pmatrix} q_{11} & \cdots & q_{1d} \\ \vdots & & \vdots \\ q_{1d} & \cdots & q_{dd} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} + 4(y_1 \cdots y_d)\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} + 2$

$$= -10\left[(x_1q_{11} + \cdots + x_dq_{1d}) \ (x_1q_{12} + \cdots + x_dq_{2d}) \ \cdots \ (x_1q_{1d} + \cdots + x_dq_{dd})\right]\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} + 4(x_1y_1 + \cdots + x_dy_d) + 2$$

$$= -10\big[(x_1x_1q_{11} + x_1x_2q_{12} + \cdots + x_1x_dq_{1d}) + (x_2x_1q_{11} + x_2x_2q_{22} + \cdots + x_2x_dq_{2d}) + \cdots + (x_dx_1q_{d1} + x_dx_2q_{d2} + \cdots + x_dx_dq_{dd})\big] + 4(x_1y_1 + \cdots + x_dy_d) + 2$$

$x_ix_j = x_jx_i$ (commutativity of x in $\mathbb{R}$)

$q_{ij} = q_{ji}$

by symmetry of Q

$$= -10\left[\sum_{i=1}^d x_i^2q_{ii} + \sum_{j=1}^d\sum_{\substack{i=1 \\ j\neq i}}^d x_ix_jq_{ij}\right] + 4\left[\sum_{i=1}^d x_iy_i\right] + 2$$

$$= -10\left[\sum_{i=1}^d x_i^2q_{ii} + 2\cdot\sum_{i<j}\sum x_ix_jq_{ij}\right] + 4\left[\sum_{i=1}^d x_iy_i\right] + 2$$

$$\Rightarrow \frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \cdots \ \frac{\partial f}{\partial x_d}\right]^T$$

$$= \begin{bmatrix} -10(2x_1q_{11} + 2x_2q_{12} + \cdots + 2x_dq_{1d}) + 4y_1 \\ -10(2x_2q_{22} + 2x_1q_{21} + 2x_2q_{23} + \cdots + 2x_dq_{2d}) + 4y_2 \\ \vdots \\ -10(2x_dq_{dd} + 2x_1q_{1d} + \cdots + 2x_{d-1}q_{d(d-1)}) + 4y_d \end{bmatrix}$$

$$= \begin{bmatrix} -20x^Tq_1 + 4y_1 \\ -20x^Tq_2 + 4y_2 \\ \vdots \\ -20x^Tq_d + 4y_d \end{bmatrix}$$

where $q_i :=$ the i^{th} col(row) of Q

5.

5. $f(x) = 8x^TQx - 2y^TQ^Tx + 6$

Q is symmetric $\Rightarrow q_{ij} = q_{ji}\ \forall i,j$

$x, y \in \mathbb{R}^d$, $Q \in \mathbb{R}^{d\times d}$

$$= 8(x_1 \cdots x_d)\begin{pmatrix} q_{11} & \cdots & q_{1d} \\ \vdots & & \vdots \\ q_{d1} & \cdots & q_{dd}\end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_d\end{pmatrix} - 2(y_1 \cdots y_d)\begin{pmatrix} q_{11} & \cdots & q_{1d} \\ \vdots & & \vdots \\ q_{1d} & \cdots & q_{dd}\end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_d\end{pmatrix} + 6$$

$$= 8\big((x_1q_{11}+\cdots+x_dq_{d1})\ (x_1q_{12}+\cdots+x_1q_{2d}) \cdots (x_1q_{1d}+\cdots+x_dq_{dd})\big)\begin{pmatrix} x_1 \\ \vdots \\ x_d\end{pmatrix}$$
$$-2\big((y_1q_{11}+\cdots+y_dq_{1d})\ (y_1q_{12}+\cdots y_dq_{2d}) \cdots (y_1q_{1d}+\cdots+y_dq_{dd})\big)\begin{pmatrix} x_1 \\ \vdots \\ x_d\end{pmatrix} + 6$$

$$= 8\big(x_1(x_1q_{11}+\cdots+x_dq_{1d}) + x_2(x_1q_{12}+\cdots+x_dq_{2d}) + \cdots + x_d(x_1q_{1d}+\cdots+x_dq_{dd})\big)$$
$$-2\big(x_1(y_1q_{11}+\cdots+y_dq_{1d}) + \cdots + x_d(y_1q_{1d}+\cdots+y_dq_{dd})\big) + 6$$

$$= 8\left(\sum_{i=1}^d x_i^2q_{ii} + \sum_{i\neq j}\sum x_ix_jq_{ij}\right)$$
$$-2\big(x_1y_1q_{11}+\cdots+x_1y_dq_{1d} + x_2y_1q_{12}+\cdots+x_2y_dq_{2d}+\cdots + x_dy_1q_{1d}+\cdots+x_dy_dq_{dd}\big) + 6$$

by symm. of Q
$$= 8\left(\sum_{i=1}^d x_i^2q_{ii} + 2\sum_{i<j}\sum x_ix_jq_{ij}\right)$$
$$-2\left(x_1\sum_{i=1}^d y_iq_{1i} + x_2\sum_{i=1}^d y_iq_{2i} + \cdots + x_d\sum_{i=1}^d y_iq_{di}\right) + 6$$

$$\Rightarrow \frac{\partial f}{\partial x} = \left[\left(16\sum_{i=1}^d x_iq_{1i} - 2\sum_{i=1}^d y_iq_{1i}\right) \cdots \left(16\sum_{i=1}^d x_iq_{di} - 2\sum_{i=1}^d y_iq_{di}\right)\right]^T$$
$$= \left[(16x^Tq_1 - 2y^Tq_1) \cdots (16x^Tq_d - 2y^Tq_d)\right]^T$$
$$= \left[(16x^T - 2y^T)q_1 \cdots (16x^T - 2y^T)q_d\right]^T$$

where q_i is the i^{th} col. of Q

$$\frac{\partial f}{\partial Q} = \begin{pmatrix} \frac{\partial f}{\partial q_{11}} & \cdots & \frac{\partial f}{\partial q_{1d}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial q_{1d}} & \cdots & \frac{\partial f}{\partial q_{dd}}\end{pmatrix} \rightsquigarrow \text{symmetric}$$

$$\begin{pmatrix}
8x_1^2 - 2x_1y_1 & 16x_1x_2 - 2x_1y_2 & \cdots & & 16x_1x_d - 2x_1y_d \\
16x_1x_2 - 2x_1y_2 & 8x_2^2 - 2x_2y_2 & & & 16x_2x_d - 2x_2y_d \\
\vdots & & \ddots & & \vdots \\
16x_1x_{(d-1)} - 2x_1y_{(d-1)} & & & & 16x_{d-1}x_d - 2x_dy_d \\
16x_1x_d - 2x_1y_d & \cdots & & 16x_{(d-1)}x_d - 2x_dy_d & 8x_d^2 - 2x_dy_d
\end{pmatrix}$$

$$\frac{\partial f}{\partial q_{ij}} = \begin{cases} 8x_i^2 - 2x_iy_i & i = j \\ 16x_ix_j - 2x_iy_j & i \neq j \end{cases}$$

6.

6. $f(x) = \|4x\|_2^2 \quad x \in \mathbb{R}^d$

$$= \left(\sum_{k=1}^{d} |4x_k|^2\right)^{1/2 \cdot 2} = \sum_{k=1}^{d} |4x_k|^2$$

$$= 16 \sum_{k=1}^{d} x_k^2$$

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_d}\right]^T = \left[32x_1 \cdots 32x_d\right]^T$$

$$= 32x$$